THE UNIVERSITY
of EDINBURGH

Characterising the role of *TLE1* in Crohn's Disease

Nidhi Sharma

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2015

Table of Contents

Tables and Figures

Declaration

I hereby declare that this thesis has been composed by me and that the work presented here is my own, unless otherwise stated, and has not been submitted for any other degree or professional qualification.

Nidhi Sharma

Acknowledgements

First and foremost I would like to thank my supervisors, Dr Elaine Nimmo and Professor Jack Satsangi. They are the most patient, kind and supportive supervisors I could ever have asked for and without their advice and continuous encouragement I would never have made it to the end. It has been an absolute privilege to work with you.

Our friendly laboratory has been a lovely place to spend the last three years, thanks to everyone who was always willing to lend a hand and offer advice. Particular thanks go to Alex Adams and Nick Kennedy for their infinite patience in teaching me how to use R and answering my statistics questions and to Helen Newbery for her experimental advice and encouragement.

To excellent proof readers, Kate, Nneka, Helen and Graham, thank you for your patience and hard work, it was much appreciated! Particular thanks to Kate for basically reading my entire thesis, to Nneka for your swiftness and Graham for all the hugs and encouragement.

A special thank you goes out to my PhD pals, Niamh, Peter, Kate, Qian, Nek-Nek and Evi, for listening to all my experimental woes and for all the fun times we had trying to forget them. Niamh thank you for your cups of tea, hugs and positivity, you are the most gifted spirt lifter I have ever met. Nek- Nek, what can I say sweet cheeks, Edinburgh just would not have been the same without your knapsack, endless brunches, Lion King sessions and N parties. Kate, you are the most supportive friend I could have asked for, thank you for counselling me almost constantly for the last year, for inspiring me with your rendition of Wuthering heights and for letting me invade your friendship group, you are wonderful and I sincerely doubt I would have survived this experience without you.

Finally, a thank you to my family, for always being ready to proof read, for not getting mad when they came to visit and I was still writing my thesis and for always being so understanding. Dad, thanks for supporting me during these last few years and for the life advice, as ever you are always spot on. Particular thanks for not letting me give up and pursue a career on X factor. Shiv, you are such an understanding and supportive soul, well done for choosing to be the better paid and understood type of doctor! Mum, cheesy as it is, you are my bestest pal, my biggest fan, my favourite cook and my inspiration to hang in there. Most of all thank you for staying up late reading me primer sequences and stopping me living off a diet of frozen chips.

Finally, to everyone else who has suffered, supported and shared their expertise with me over the last four years, THANK YOU!

Abbreviations

AGR2	Anterior gradient 2
APAF1	Apoptotic protease activating factor 1
APC	Antigen presenting cell
ATG16L1	Autophagy related 16-like 1
b(HLH)	Basic helix loop helix
BiP	Binding immunoglobulin protein
CARD	Caspase recruitment domain
CCND1	Cyclin dependant kinase 1
CD	Crohn's disease
CDC2	Cyclin dependant kinase 2
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CK2	Caesin kinase 2
CMV	Cyclomegalovirus
Co-IP	Co-immunoprecipitation
CRC	Colorectal cancer
DAMP	Damage associated molecular pattern
DC	Dendritic cell
dH_2O	Distilled water

DMEM	Dulbecco's modified eagles medium
DMSO	Dimethyl sulfoxide
dNTPs	Deoxynucleotides
DSS	Dextran sulphate sodium
EDTA	Ethylinediamenetetraacetic
EH1	Engrailed homology 1
ER	Endoplasmic reticulum
EST	Expressed sequence tag
FCS	Fetal calf serum
FDR	False discovery rate
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GI	Gastrointestinal
GWA	Genome wide association
H_2O	Water
HC	Healthy control
HDAC	Histone deacetylase
HES	Hairy enhancer of split
HRP	Horseradish peroxidase
HSP90	Heat shock protein 90
IBD	Inflammatory Bowel Disease
IHC	Immunohistochemistry

IKK	IκB kinase
IL23	Interleukin 23
IL23R	Interleukin 23 receptor
IRGM1	Immunity related GTPase family M
JAK2	Janus kinase 2
LC3	Light chain 3
LDL	Low density lipoprotein
LPS	Lipopolysaccharide
LRR	Leucine rich repeat
LRRK2	Leucine rich repeat kinase 2
MDP	Muramyl dipeptide
MEKK4	Map kinase kinase 4
MUC2	Mucin 2
NFκB	Nuclear factor κB
NICD	Notch intracellular domain
NLR	NOD like receptor
NOD2	Nucleotide oligomerisation domain 2
PAMP	Pattern associated molecular pattern
PBS	Phosphate Buffered Saline
PC	Paneth cell
PCR	Polymerase chain reaction

PGN	Peptidoglycan
PRR	Pattern recognition receptor
RIOK1	RIO Kinase 1
RIPK2	Receptor interacting serine/threonine kinase 2
RNA	Ribonucleic acid
RT qPCR	Real time quantitative PCR
RUNX	Runt related transcription factor
SDHA	Succinate dehydrogenase
SDS	Sodium dodecyl sulphate
SGPL1	Sphingosine phosphate lyase 1
SLC	Solute carrier protein
SMURF1	SMAD specific E£ ubiquitin protein ligase 1
SNP	Single nucleotide polymorphism
STAT3	Signal transducer and activator of transcription 3
TAK1	Transforming growth factor -beta activated kinase
TBP	Tata binding protein
TCF/LEF	T cell factor/Lymphoid enhancing factor
TF	Transcription factor
TGFβ	Transforming growth factor β
TLE1	Transducin-like enhance or Split
TLR	Toll-like receptor

TNFα	Tumour necrosis factor α
TSL	Transcript support level
TUSC3	Tumour suppressor candidate 3
UBC	Ubiquitin C
UC	Ulcerative Colitis
UPR	Unfolded protein response
WH	Winged helix
WNT	Wingless
XBP1	X box binding protein 1

Abstract

The inflammatory bowel diseases (IBD) are chronic, relapsing and remitting diseases of the gastrointestinal tract. There are two main types of IBD: Crohn's disease (CD) and ulcerative colitis (UC). The prevalence of IBD is highest in the western world, approximately 100-200 people per 100,000 are affected. In recent years there has been a marked increase in the incidence of CD and UC, in both adults and children (Henderson et al., 2012; Molodecky et al., 2012). This is particularly relevant in Scotland where recent research shows that there has been a 79% increase in the number of cases of paediatric IBD since the 1990's (Henderson et al., 2012).

A yeast 2 hybrid screen identified TLE1as an interacting partner of the known CD susceptibility gene; *Nucleotide- binding oligomerisation protein 2* (*Nod2*). An initial genome wide association study (GWAS) also found an association between the rs6559629 SNP, located in *Tle1* and ileal CD ($p = 3.1 \times 10^{-5}$) and showed that carriage of the *Tle1* risk allele increases the effects of *Nod2* mutations in CD. TLE1 functions as a transcriptional co repressor in a variety of different cellular and developmental pathways

The work presented in this thesis investigates the potential role of *TLE1* in CD. This has been approached using four different strategies: sequencing *TLE1* in CD patients and controls, analysing the effects of knocking down *TLE1* on genome wide expression, investigating whether the known IBD susceptibility protein XBP1 binds to a predicted binding site in *TLE1* and investigating TLE1 levels and localisation in human intestinal samples from CD patients and controls

Sequencing *TLE1* exons and introns 15/16 and 16/17 in a Scottish cohort of 24 CD patients and healthy controls identified a number of potentially pathogenic exonic and intronic SNPs. Two exonic SNPs and thirteen intronic SNPs were identified and these were further investigated in larger Scottish (203 CD cases, 190 HC) and European cohorts (6,333 CD cases and 15,056 HC) but were not present at statistically significantly different frequencies.

Secondly, the effects of *TLE1* knock down on genome wide expression were analysed using an Illumina HT12 expression chip. The results showed that *TLE1* knock down significantly altered expression of 19 loci (Bonferroni) and 526 loci (FDR). Four of the 19 Bonferroni significant loci are potentially involved in CD: *RIOK1* ($p=4.3\times10^{-3}$), *SGPL1* ($p=4.3\times10^{-3}$), *TUSC3* ($p=1.8\times10^{-2}$) and *CCND1* ($p=2.7\times10^{-3}$). Furthermore, expression of *SGPL1* and *RIOK1* were shown to be differentially expressed at the mRNA level between inflamed patients and controls.

The third approach investigates a predicted binding site for the known IBD susceptibility gene, *XBP1* in *TLE1* which was identified using the Haploreg program. This work shows, using chromatin immunoprecipitation, that exogenous *XBP1* does not appear to bind to this predicted binding site.

Finally, TLE1 expression was analysed in human intestinal resection samples from patients of known *NOD2* status. This work shows that TLE1 and NOD2 are expressed in Paneth cells, however TLE1 expression is not altered in patients carrying CD associated *NOD2* variants.

In this work *TLE1* sequence, expression and potential interacting proteins have been analysed. The results presented suggests multiple mechanisms by which *TLE1* may be influencing susceptibility to CD including: the unfolded protein response (*TUSC3*), S1P signalling and ribosome biogenesis. They also implicate TLE1 in Paneth cell function alongside NOD2. The exact means by which TLE1 may play a role in IBD pathogenesis has yet to be fully elucidate

Chapter 1: Introduction

1.1 Inflammatory Bowel Disease

The inflammatory bowel diseases (IBD) are chronic, relapsing and remitting diseases of the gastrointestinal tract. In IBD, the most commonly affected regions of the tract are the colon and ileum. There are two main types of IBD: Crohn's disease (CD) and Ulcerative colitis (UC). The symptoms of these disorders are similar and include diarrhoea, abdominal pain, gastrointestinal bleeding, weight loss, malnutrition and fatigue. Furthermore, UC sufferers have 60% increased risk of developing colorectal cancer when compared to that of the general population (Herrinton et al., 2012).The key differences between UC and CD are the region of the GI tract affected and the degree of inflammation. Clinically, CD can be characterised by discontinuous inflammation in any part of the gastrointestinal tract from the mouth to anus. However, CD most commonly affects the lower part of the small intestine called the ileum, however, approximately 5% of cases affect the mouth. This inflammation is transmural; it can affect any layer of tissue. In UC there is usually continuous distribution of inflammation which begins in the rectum and can extend **to** the caecum. The inflammation which usually begins in the lamina propria layer of tissue and is limited to the mucosal and submucosal layers of the intestinal wall. In a proportion of cases with colonic inflammatory bowel disease patients show characteristics of both UC and CD, these patients are diagnosed with IBD unclassified (IBD-U).

1.2 Structure of the gastrointestinal (GI) tract

The gastrointestinal tract contains the organs responsible for the ingestion of food, the digestion of food and the excretion of any waste material. The upper GI tract is comprised of the oesophagus, stomach and the first part of the small intestine, the duodenum. The lower section of the tract comprises the rest of the small intestine (jejunum and ileum) and the large intestine (colon, caecum and rectum).

The process of digestion begins with the mouth and salivary glands which orchestrate the mechanical and enzymatic breakdown of food. This process is continued in the stomach, where the food is further broken down in the presence of acids and enzymes. The partially digested food then passes into the small intestine which contains

digestive enzymes secreted by the pancreas. The small intestine has three sections: the duodenum, the jejunum and the ileum; these are responsible for the final stages of digestion and the absorption of nutrients. To aid absorption of nutrients, the surface of the small intestine is comprised of many crypts which provide increased surface area. Following digestion in the small intestine, the remaining material passes through the ileo-colonic valve into the colon allowing absorption of water and vitamins, leftover waste material is stored in the rectum prior to excretion.

Histologically, the GI tract is composed of four layers (from inside to outside): the mucosa, submucosa, muscularis externa and the adventitia. These layers are evident throughout the GI tract, however their composition varies. In the small intestine, the inner mucosal layer consists of three sub-layers the inner epithelial layer, the lamina propria and the muscularis mucosa.

The inner epithelial layer acts as a barrier to defend against bacterial invasion, whilst facilitating efficient absorption of nutrients and fluids and excretion of waste material. In the small intestine the epithelial layer consists of absorptive (epithelial cells) and secretory cells (endocrine cells, goblet cells and Paneth cells). Endocrine cells produce hormones that are release into the bloodstream through the copious amounts of blood vessels in the mucosa. Goblet cells produce glycosylated proteins called mucins. These proteins form a mucus which is secreted by the goblet cells and functions to protect the epithelial cells from enzymatic and bacterial damage. Paneth cells are involved in: bacterial sensing, upon exposure to bacteria or bacterial antigens, they release antimicrobial peptides including lysozyme and α defensins. The production of these peptides helps to retain the integrity of the epithelial barrier and prevent bacteria from entering the intestinal crypts. Microfold cells (MC) and dendritic cells (DC) also form part of the epithelial cell layer and along with Paneth cells play a key role in the intestinal immune system. In response to the presence of bacteria or defects in the epithelial barrier DC's are activate and migrate to the lymph nodes where they induce the differentiation of naïve T cells into T effector and T regulatory cells. The lamina propria is a thin layer of connective tissue containing patches of lymphoid tissue which

produce macrophages and lymphocytes to protect the GI tract and a vascular system which supports the epithelial cell layer. The muscularis mucosa is a thin layer of smooth muscle which supports movement of the mucosa.

The sub mucosa is a connective tissue layer containing larger blood vessels as well as nerve, lymph and glandular tissue. The muscularis externa is composed of two layers of muscle, a circular layer and a longitudinal layer. The function of this layer is to enable peristalsis, which pushes food through the gut. The outermost layer of tissue is an outer layer of connective tissue, the serosa.

1.2.1 Incidence of IBD

The prevalence of IBD is highest in the western world, approximately 100-200 people per 100,000 are affected. In recent years there has been a marked increase in the incidence of CD and UC, in both adults and children (Henderson et al., 2012; Molodecky et al., 2012). This is particularly relevant in Scotland as recent research shows that there has been a 79% increase in the number of cases of paediatric IBD since the 1990's (Henderson et al., 2012). The cause of this increase is not known, however a number of potential environmental risk factors have been identified, namely smoking, diet and exposure to bacteria.

1.2.2 Pathogenesis

Both UC and CD are polygenic, complex diseases that are thought to be caused by a combination of genetic and environmental susceptibility factors.

The precise aetiology of IBD has yet to be fully elucidated. It is generally thought to arise from a dysregulated response to luminal bacteria in the intestinal epithelium in genetically susceptible individuals. In line with this, current treatment is limited to dietary supplementation, administering of drugs such as immunosuppressant's and anti-inflammatory drugs and surgery. These treatments do not provide a cure and have adverse side effects, for example the administration of immunosuppressant's has is

known to increase the risk of IBD patients developing skin cancer (Long et al., 2012) .

IBD is a lifelong disease, requiring expensive medication and multiple surgeries, and therefore represents a significant burden on health services. Research into the molecular genetic basis underlying the cause of this disease is essential in order to develop more effective treatments.

1.3 Environmental risk factors for IBD

Many different environmental factors have been suggested to increase risk of developing IBD including smoking, oral contraception, diet, breast feeding, drugs, geographical and social status, microbial agents, intestinal permeability and appendectomy , (reviewed in (Molodecky & Kaplan, 2010)). However, it is difficult to quantify the individual contribution of each of these factors to disease risk. The most widely accepted environmental IBD risk factor is smoking; it has been shown to affect both IBD incidence and history. Interestingly, smoking is a risk factor for developing CD but a protective factor for UC (M. C. Aldhous & Satsangi, 2010).

1.4 Genetics of IBD

1.4.1 Heritability

Epidemiological data provide evidence suggesting that genetics contributes significantly to IBD susceptibility (Molodecky et al., 2012). There is variation in incidence and prevalence of IBD in different ethnic groups, for example Ashkenazi Jews have a two to four fold higher incidence and prevalence than non-Jewish Caucasians (Roth, Petersen, McElree, Feldman, & Rotter, 1989; H. Yang et al., 1993).

Twin and familial aggregation studies have shown that IBD has a strong genetic component, between 5-20% of patients have a family history of IBD which is more apparent in CD than UC (Orholm & et al, 1991). There is increased concordance for IBD in monozygotic twins than dizygotic twins, and this is higher is CD (20-50% in MZ vs 0- 7% in DZ) than in UC (14-19% in MZ vs 0- 5% in DZ) (Halfvarson & et al,

2003; Orholm, Binder, Rasmussen, Kyvik, & Hospital, 2000; Thompson, Driscoll, Pounder, & Wakefield, 1996; Tysk & Lindberg, 1988).

It is important to note that overall estimates of heritability in IBD are subject to controversy as they do not account for interactions between genes, environmental risk factors or interactions between genes and the environment.

1.4.2 Approaches

Initial approaches to the investigation of the heritability of IBD involved linkage analysis which led to the led to the identification of two IBD risk loci in 1996: IBD1 (Chromosome 16) and IBD9 (Chromosome 3). Further dissection of the IBD1 locus led to the identification of the first CD susceptibility gene in 2001: *NOD2* (Hugot et al., 2001; Ogura et al, 2001; Peter & Muddassar, 2001). Three low frequency polymorphisms in the coding region of *NOD2* were discovered (R707W, G908R and L1007fs) (Figure 4). It has since been shown that these mutations, in their homozygous or compound heterozygous states, result in a 20-40 fold increased risk of developing CD (Bonen & Cho, 2003). These mutations and the function of *NOD2* are central to this project and are further discussed in section 1.4.6. Aside from the discovery of *NOD2*, linkage studies found very few replicable IBD susceptibility loci. This may have been due to the relatively small sample sizes used in these studies (usually between 100 and 400) leading to false findings. This information suggested that the genetic architecture of IBD was likely to be much more complex than that of Mendelian diseases and was unlikely to be caused by a defect in a single gene. It was postulated that complex diseases, such as IBD could be caused by a combination of large numbers of common variants of small effect size.

Advances in technology led to the advent of genome wide association studies (GWAS). These are large scale studies, typically involving many thousands of cases and controls. These studies use a hypothesis free approach and involve comparing the frequency of single nucleotide polymorphism (SNP) alleles across the genome in cases and controls. This allows identification of SNPs that are associated with a disease, SNPs tagging a region of DNA implicated in disease pathogenesis. GWA studies have

led to the identification to 201 susceptibility loci (J. Z. Liu et al., 2015). Of these 201 loci, the majority (148) are implicated in both CD and UC, 30 are only relevant to CD and 23 are specific to UC. Together these loci account for 13.1% and 8.2% of the genetic heritability of CD and UC, respectively (J. Z. Liu et al., 2015). These loci are involved in many biological pathways including: innate pattern recognition (NOD2/ CARD15), differentiation of Th17-lymphocytes (IL23, IL23R), autophagy (ATG16L1) and the unfolded protein response (XBP1). GWA studies have not replicated many of the loci identified by linkage studies. Although as previously discussed this may be due to the limited cohort sizes of linkage studies, it is important to note that unlike linkage studies, the number of patients with a family history of IBD in GWAS cohorts is low. Additionally, mutations in *NOD2,* which is the strongest CD risk gene identified to date, were identified by linkage studies and have been replicated in GWA studies (Hugot et al., 2001; Jostins et al., 2012; Ogura et al, 2001).

1.4.3 The differentiation of Th17 lymphocytes in IBD pathogenesis

Th17 cells are a subset of thymus derived lymphocytes that express the proinflammatory cytokine Interleukin 17 (IL 17). These cells are positive for CD4, a glycoprotein present on the cell membrane of certain subtypes of immune cells. They are characterised by the synthesis of the cytokines: IL 17A, IL 21A, IL 22 and the expression of the receptors: IL23R and CCR6 (Wilson et al., 2007). Th17 cells function as part of the adaptive immune response. Upon exposure to antigens, naïve Th17 cells differentiate in order to perform a number of effector functions i.e. production of proinflammatory cytokines and anti apoptotic proteins (Figure 1). This differentiation process is induced by the cytokine IL 23. IL 23 interacts with the heterodimeric IL 23 receptor (IL23R). Upon binding of IL 23 to IL 23R, JAK2 is autophosphorylated and STAT3 forms an active homodimer and translocates into the nucleus to initiate transcription.

GWAS studies have identified variants in the IL 23 receptor as associated with CD (Duerr et al., 2006). Duerr *et al* showed that the Arg381Gln substitution in *IL23R* has a protective effect in CD. More recent work has used deep sequencing of CD patients to identify rare variants in IL23R with a protective effect (Balzola, Bernstein, Ho, & Russell, 2012).

The cytokines IL17A, IL22 and IL 26 show increased expression in both intestinal tissue and serum of IBD patients (Andoh et al., 2005; Fujino et al., 2003). Furthermore, a 20 fold increase is seen in the number of active Th17 cells in gut resections taken from CD patients when compared to controls. These Th17 cells have been shown to promote inflammation in colonic cell lines, treatment of cells with supernatant from Th17 cells (containing cytokines released by Th17 cells) leads to increased expression of chemokines (CXCL, CXCL8, CCL20) that are involved in maintenance of inflammation in CD patients (Kleinschek et al., 2009).

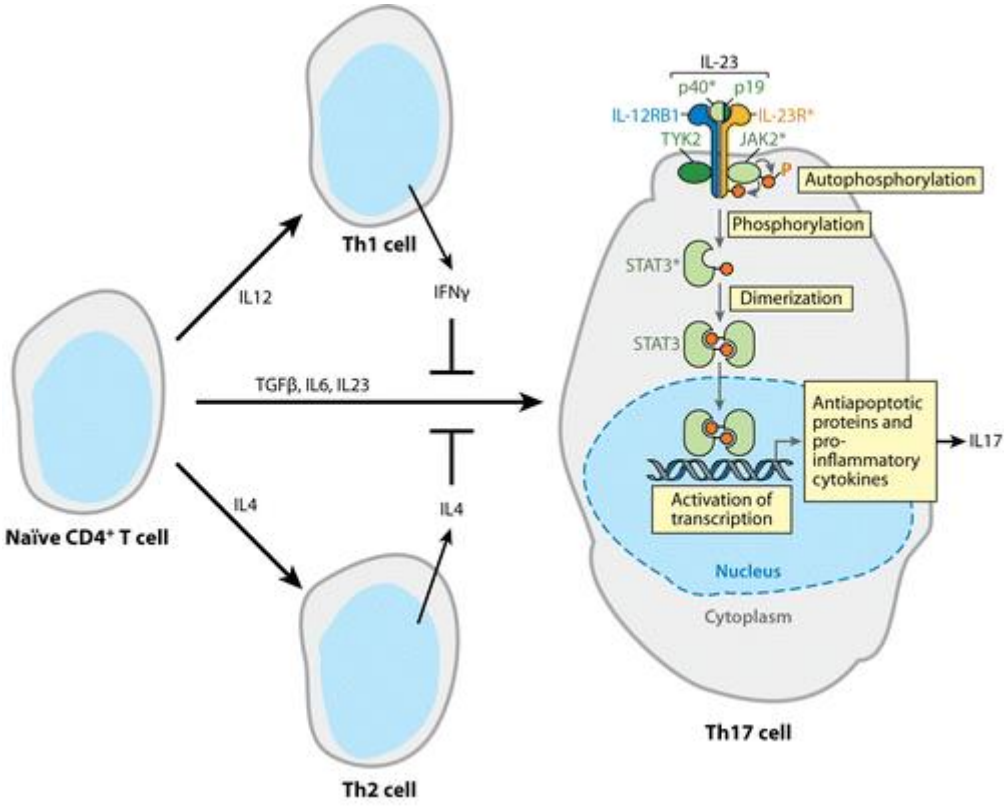


Figure 1 T helper (Th) cell differentiation

Adapted from (Van Limbergen, Wilson, & Satsangi, 2009). Naïve T cells differentiate into Th1, Th2 or Th17 cells depending on cytokine stimulation. Th17 cells are induced by TGFβ, IL6 and IL23. IL23 binds to the IL23 receptor on the surface of Th17 cells, causing JAK2 to be phosphorylated, STAT3 to homodimerise and translocate into the nucleus. STAT3 initiates transcription of pro inflammatory cytokines and anti- apoptotic proteins.

1.4.4 The Unfolded protein response in IBD pathogenesis

The rough endoplasmic reticulum (ER) is the eukaryotic organelle responsible for the synthesis, folding and post translational modification of proteins. When the burden on the ER is too high, unfolded/misfolded proteins accumulate in the ER, this situation is defined as ER stress. The accumulation of unfolded/misfolded proteins initiates the unfolded protein response (UPR) which functions to resolve ER stress or induce apoptosis. *X box protein 1* (*XBP1*) encodes a transcription factor that functions as part of the unfolded protein response. The role of XBP1 in IBD is discussed in detail in chapter 3. In brief, both common SNPs in the *XBP1* region and four rare variants have been associated with IBD (Kaser et al., 2008). Additionally, XBP1 null mice show

increased ER stress, develop spontaneous enteritis and have an exaggerated response to pro inflammatory cytokines (IBD inducers) e.g. TNFα and Flagellin in the epithelium. They also have an increased susceptibility to colitis as a result of Paneth cell absence and malformation (Kaser et al., 2008).

Anterior gradient 2 (*AGR2*) is a protein disulphide isomerase which aids formation of disulphide bonds in the endoplasmic reticulum (ER). AGR2 aids formation of disulphide bonds in the cysteine rich Mucin 2 (MUC2) protein. MUC2 is a key component of the mucus secreted by goblet cells. Mice lacking *Agr2* show decreased MUC2 levels and increased expression of the ER stress markers *Xbp1(sp)* and *BiP*. AGR2 is highly expressed in the ileum and colon and shows decreased expression levels in gut biopsies from UC patients when compared to healthy controls (W. Zheng et al., 2006). Additionally, FOXA1 and FOXA2 regulate goblet cell differentiation and have been shown to activate the *AGR2* promoter region using luciferase reporter assays (W. Zheng et al., 2006).

1.4.5 IBD and Autophagy

Autophagy describes the evolutionarily conserved process by which cellular components are degraded in lysosomes. It is mediated by the autophagosome which consumes portions of cytoplasm. The process of autophagy begins with the formation of a phagosome which is (double membraned structure) forming around molecules that are to be degraded. This phagosome then expands, elongates and matures, eventually fusing with lysosomes resulting in the degradation of components encompassed by the autophagosome. *Autophagy related gene like-1* (*ATG16L1*) encodes an autophagy protein involved in the formation and expansion of the autophagosome. During autophagosome formation, ATG16L1 forms a complex with other autophagy proteins (ATG5 and ATG 12) to enable conversion (lipidation) of the microtubule associated light chain protein (LC3) to LC3-II. LC3-II is a marker of autophagy and is located on the membrane of the mature autophagosome. The ATG16L1 T300A variant has been associated with CD, this variant causes a threonine to alanine substitution (Hampe et al., 2007)

In immune cells, autophagy is important for degradation of pathogens and initiation of the immune response. It is up regulated during intracellular infection, oxidative stress and upon accumulation of misfolded proteins.

A 20Kb deletion upstream of the *Immunity related GTPase M* (*IRGM*) locus has also been associated with CD. Similarly to ATG16L1, *IRGM* has been implicated in the conversion of LC3-1 to LC-II. Human macrophages treated with *IRGM* siRNA do not form LC3-II in the presence of the autophagy inducer rapamycin (Singh, Davis, Taylor, & Deretic, 2006). Additionally, *IRGM* is involved in bacterial resistance. Increased bacterial survival is seen in human cells transfected with *IRGM* siRNA when compared to control transfections (Singh et al., 2006). *Igrm1* knockout mice exhibit increased acute ileal and colonic inflammation in response to DSS treatment when compared to controls. Furthermore, Paneth cells from these mice showed altered morphology of secretory granules and reduced expression of antimicrobial peptides including α defensins (B. Liu et al., 2013).

Other autophagy proteins have also been associated with IBD including: *Leucine repeat kinase 2* (*LRRK2*) and *SMAD Specific E3 Ubiquitin Protein Ligase 1* (*SMURF1*) (Jostins et al., 2012).

1.4.6 NOD2 (Nucleotide Oligomerisation Domain 2)

Structure and Function

Nucleotide oligomerisation protein 2 (*NOD2*) encodes a member of the Nod like Receptor (NLR) protein family. *NOD2* functions as an intracellular pattern recognition receptor that plays a crucial role in innate immunity and inflammation. It consists of three protein domains, two caspase recruitment domains (CARD) at the N terminus, a nucleotide binding domain (NBD) and a C terminal leucine rich repeat region (LRR) (Figure 2). The LRR is thought to be responsible for ligand detection, it recognises bacterial components. *NOD2* has been shown to bind to the bacterial cell wall

component, Muramyl Dipeptide (MDP) , upon binding a downstream signalling cascade initiates pro inflammatory, immune and anti- microbial signalling pathways (Grimes, Ariyananda, Melnyk, & O'Shea, 2012). It is not known whether accessory proteins are necessary for the NOD2 –MDP interaction as is the case with Toll-like receptor 4 (TLR4) which forms a heterodimer with myeloid differentiation factor- 2 (MD-2) to recognise lipopolysaccharide (LPS) (Park et al., 2009). The nucleotide binding domain of *NOD2* contains winged helix (WH) and helix domains (HD) which play an important role in NOD2 oligomerisation. X ray crystallography of apoptosis protease activator 1 (*APAF1*) a member of the NOD protein family suggests that when inactive, NOD2 is in a monomeric state where the WH domain is bound to ADP and the protein has a closed conformation (Riedl, Li, Chao, Schwarzenbacher, & Shi, 2005). When MDP binds to the LRR, NOD2 protein conformation changes, ADP-ATP exchange occurs, NOD2 oligomerises and downstream signalling is initiated (Riedl et al., 2005). The suggested functions of the LRR and NOD domains are supported by analysis of mutations in these regions which lead to ineffective bacterial sensing (and therefore downstream NFκB signalling) and lack of NOD2 inhibition (constitutive NFκB activation) respectively (Tanabe et al., 2004).

Figure 2 Structure of NOD2

Adapted from (Rivas et al., 2012). NOD2 consists of two N terminal caspase recruitment (CARD) domains a nucleotide binding domain (NBD) and a leucine rich repeat region (LRR). The CARD domains are where protein interactions occur, for example RIP2 binds to this region. The NBD is where ATP binds and is metabolised and where NOD2 oligomerisation occurs. The LRR domain at the carboxyl terminal is where ligands including MDP and LPS are recognised.

NOD2 functions as part of the innate immune system

The innate immune system is the human body's first line of defence upon exposure to pathogens. It is a rapidly induced, evolutionarily conserved, response to microbial invasion. The human gut is constantly exposed to both harmful and harmless microbes. An overactive immune response is thought to be causal in IBD. The key functions of the innate immune system are: recognition of microbes, regulation of inflammation, and activation of the adaptive immune response. Pattern recognition receptors (PRRs) are responsible for the recognition of microbes. PRRs include the Toll- like family of receptors (TLRs) and the Nucleotide oligomerisation domain- like receptors (NODs). These PRRs recognise components of microbes called Pathogen- associated microbial patterns (PAMPs) which include bacterial/viral DNA, bacterial cell wall components, and flagella proteins. PRRs have also been shown to recognise damage-associated molecular patterns (DAMP's), these are molecules that are exposed upon destruction of host membranes and are released as a result of necrosis.

NOD2 and RIPK2

NOD2 is activated by the binding of the ligand MDP in the cytoplasm. MDP enters the cytoplasm through three mechanisms: interaction with solute carrier proteins (SLCs), invasive bacteria or cellular absorption of bacterial membrane vesicles (reviewed in (Boyle, Parkhouse, & Monie, 2014)) . Activated NOD2 oligomerises and interacts with receptor –interacting serine /threonine protein kinase 2 (RIPK2) via its CARD domains. Both CARD domains are essential for the NOD2-RIPK2 interaction and downstream NFκB signalling (Yasunori Ogura et al., 2001). RIPK2 is necessary for downstream NFκB and MAPK signalling. Embryonic fibroblasts derived from RIPK2 null mice show diminished NFκB signalling when transfected with a NOD2 over expression construct whereas cells derived from wild type mice initiate NFκB signalling when NOD2 is over expressed (K. Kobayashi et al., 2002). In unstimulated cells, where NOD2 is inactive, RIPK2 is bound by MAP Kinase Kinase 4 (MEKK4), activated NOD2 competes for RIPK2 resulting in formation of a RIPK2:NOD2 complex.

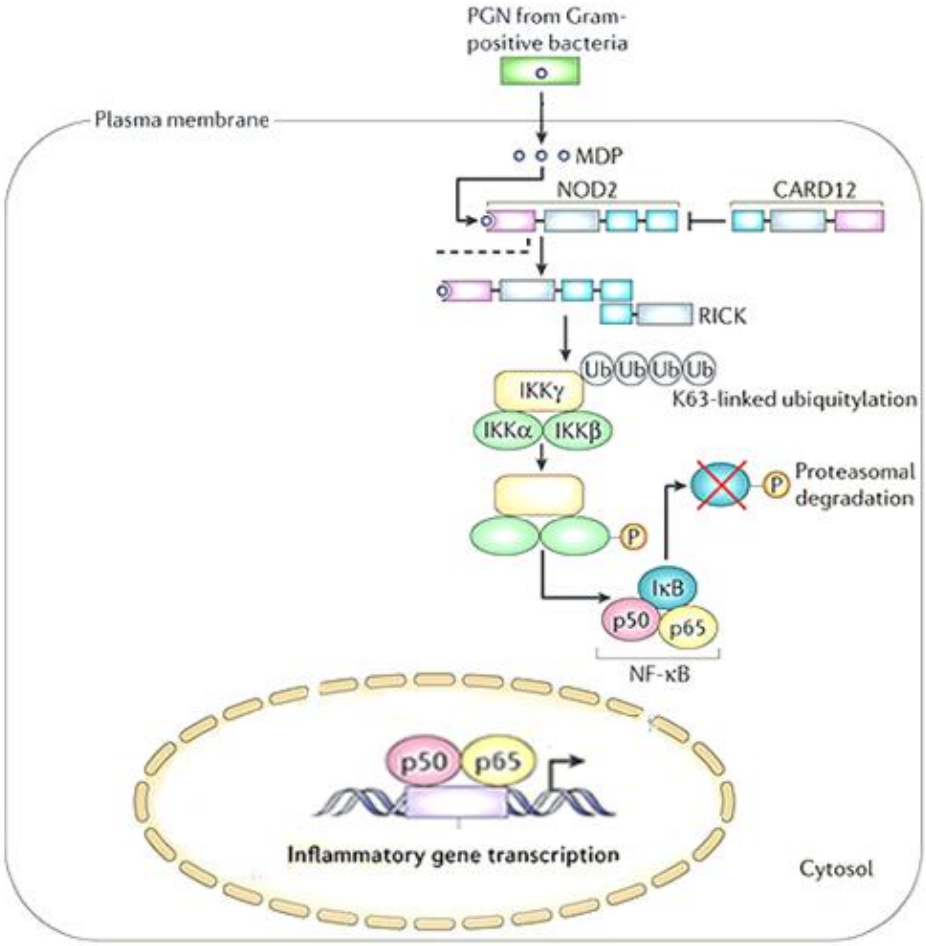


Figure 3 NOD2 in the NFκB signalling pathway

Figure adapted from (Strober, Murray, Kitani, & Watanabe, 2006). Peptidoglycan (PGN) from Gram positive bacteria enters the cell. Muramyl dipeptide (MDP), a component of PGN is recognised by NOD2. NOD2 is subsequently activated, it oligomerises and interacts with the kinase, RIPK2 (also known as RICK). Downstream NFκB signaling is initiated, the IKK complex is ubiquitinated leading to proteosomal degradation of the NFκB inhibitor IKK. An active NFκB heterodimer (p50/p65) is then able to translocate into the nucleus and initiate transcription.

NOD2 in the NFκB signalling pathway

The NFκB pathway is a pro inflammatory signaling pathway that is initiated by exposure to pro inflammatory cytokines, viruses, bacterial exposure and DNA damage.

The NFκB proteins are a family of nuclear transcription factors: RELA, RELB, c-REL, p50 and p52 which are well conserved and functionally related, particularly in the N terminal region which contains sequence necessary for nuclear localisation and dimerisation. There are two pathways by which the NFκB pathway is activated, the canonical pathway and the alternative pathway. The canonical pathway is activated by pro inflammatory cytokines such as TNFα whereas the alternative pathway is initiated by other cytokines such as RANKL and TNFSF11. The difference between these two pathways is that the canonical pathway leads to activation of NFκB RelB/p52 heterodimers whereas the alternative pathway activates the RelA and c-Rel complexes.

The general post activation mechanism of both the canonical and alternative NFκB signaling pathways is similar. When inactive, the NFκB dimers are bound by the IκB (Inhibitors of NFκB) proteins, their nuclear localization signal is not exposed and they are retained in the cytoplasm. Stimulation of the NFκB pathway leads to dissociation of NFκB dimers from their inhibitors, the IκB proteins. This dissociation is mediated by the IκB kinase (IKK) complex. This complex consists of catalytic components (IKKα, IKKβ, and IKKγ) and adaptor proteins which together phosphorylate the IκB inhibitory proteins, targeting them for proteasome mediated degradation. Active NFκB dimers can then translocate into the nucleus where they activate a range of genes involved in the immune response, cell cycle checkpoints, NFκB inhibition (negative feedback loop) and anti-apoptotic genes (Figure 3).

NOD2 activates the NFκB pathway, this activation occurs via the IKK complex. Ogura et al showed that HEK293T cells activated the NFκB signaling pathways using a luciferase assay, they also showed that decreased NFκB activation in HEK293T cells transfected with dominant negative forms of IKKα, IKKβ or IKKγ or IκB (Yasunori Ogura et al., 2001). Additionally, RIPK2 is also necessary for NOD2 dependent activation of the IKK complex. The activated NOD2:RIPK2 complex is ubiquitinated which leads to recruitment of Transforming growth factor β activated kinase 1 (TAK1). TAK1 interacts with the kinase domain of RIPK2 and the IKK complex binds to the intermediate domain of RIPK2. The IKKγ subunit is subsequently ubiquitinated

and by TAK1 and undergoes degradation allowing phosphorylation of the NFκB inhibitor, IκB by IKKα and IKKβ. This leads to release of the p65 NFκB subunit allowing translocation into the nucleus and activation of transcription.

NOD2 levels are tightly controlled, and MDP stimulation of NOD2 initiates a negative feedback loop whereby heat shock protein 90 (HSP90) dissociates from NOD2 which leads to ubiquitination and proteasome mediated degradation of NOD2 (K.-H. Lee, Biswas, Liu, & Kobayashi, 2012).

NOD2 Expression patterns

In line with its function as a PRR, NOD2 has been shown to be located in the plasma membrane and cytoplasm, where it can come into contact with bacterial products such as MDP i.e. antigen presenting cells (APCs). This is supported by NOD2 expression in macrophages and dendritic cells but not in T or B cells of the immune system (Gutierrez et al., 2002). A number of chemicals including TNFα and IFNγ activate NOD2 and increase the response of NOD2 to MDP (Rosenstiel et al., 2003). In the intestine, NOD2 expression was first detected in Paneth cells and was originally thought to be exclusively expressed in these cells (Y Ogura et al., 2003). The absence of a functioning NOD2 antibody has meant that a majority of work has looked at mRNA expression levels of *NOD2*, however recent advances suggest that NOD2 is also expressed at the protein level in intestinal epithelial cells and monocyte derived cells (dendritic cells) of the immune system (Barnich, Aguirre, Reinecker, Xavier, & Podolsky, 2005; Hu & Peter, 2013a). In dendritic cells *NOD2* expression is upregulated upon exposure to viruses (Hu & Peter, 2013b). Additionally in mice, NOD2 expression has been detected in the stem cells of the intestinal crypts (Nigro, Rossi, Commere, Jay, & Sansonetti, 2014).

Recent work indicates that NOD2 is able to shuffle to the nucleus (Barnich, Aguirre, et al., 2005; Kufer, Kremmer, Banks, & Philpott, 2006; Zurek et al., 2012). NOD2 has been detected in the nucleus in HeLa cells overexpressing NOD2. There is a 6 fold

increase in the concentration of nuclear NOD2 in cells treated with chemicals that block nuclear export (Zurek et al., 2012).

NOD2 in human disease

NOD2 is a highly conserved gene located on chromosome 16. Three mutations in *NOD2* were the first genetic risk factors to be associated with CD, particularly ileal CD (Economou, Trikalinos, Loizou, Tsianos, & Ioannidis, 2004; Hugot et al., 2001; Ogura et al, 2001). These were: R702W (missense), G908R (missense) and L1007fsinsC (frame shift) all of which affect the LRR region of the NOD2 proteins (Figure 4) (Hugot et al., 2001; Ogura et al, 2001). The R702W and G908R are point mutations leading to amino acid substitutions, arginine to tryptophan and glycine to arginine respectively. The L1007fsinsC is an insertion leading to a premature stop codon and a 1007 amino acid truncated NOD2 protein. Studies have shown that the presence of one of these NOD2 risk alleles confers a 2-4 fold increased risk whereas two risk alleles increases the risk of CD development by between 20 and 40 fold (reviewed in (Bonen & Cho, 2003)). These mutations are the strongest CD risk factors that have been identified to date and provide important insights into the pathogenesis of CD.

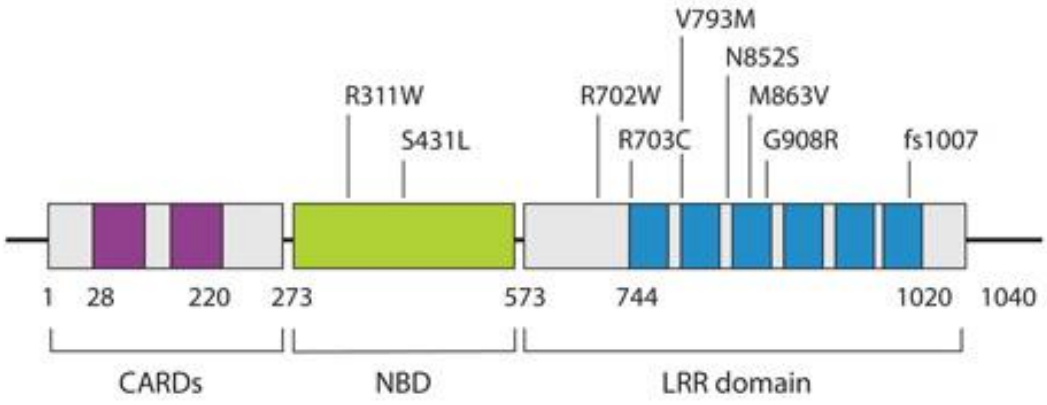


Figure 4 NOD2 mutations associated with Crohn's disease (adapted from (Rivas et al., 2012)

Schematic diagram of the NOD2 protein, the three most common CD mutations are in the LRR region (R702W, G908Rand L1007fsInsC (fs1007)). Six additional low frequency variants found in CD patients are also shown as identified by deep sequencing analysis conducted by (Rivas et al., 2012).

In addition to the more common *NOD2* variants discussed, Hugot et al identified over 30 rare *NOD2* polymorphisms and recent deep sequencing has revealed six low frequency *NOD2* risk variants (R311W, S431L, R703C, V793M, N852S and M863V) in CD patients. These six variants are not in linkage disequilibrium with the R702W, G908R or L1007fsInsC variants. Aside from the L1007fsinsC mutation, all other variants result in amino acid substitutions. The N852S, S431L, R702W, G908R and fs1007 all affect downstream NFκB signalling, in the fs1007 mutant signalling is abolished (Inohara et al., 2003; Rivas et al., 2012). The S431L and fs1007 mutant forms of NOD2 do not localise to the plasma membrane in HEK293 cells, they are retained in the cytoplasm, the N852S mutant localises similarly to wild type NOD2 in the plasma membrane (Rivas et al., 2012).

Prevalence of mutations

Although the R702W, G908R and L1007fsinsC variants are the strongest genetic risk factor for CD found to date and they are only found in approximately 36 percent of European patients (Cuthbert et al., 2002). Approximately 8 to 17 percent of CD patients have two *NOD2* risk alleles, compared to approximately 1 percent in Caucasian healthy controls. In the western population between 30 and 50 percent of people carry one NOD2 risk allele. The frequency of these mutant alleles varies dramatically across different populations, they are not found in Asian populations and are more frequent in southern Europeans than northern Europeans (Hugot et al, 2007). Additionally they are only found in approximately eleven percent of Scottish patients, suggesting that they are not attributable for the higher incidence of CD in Scottish populations (Arnott et al., 2004).

The exact mechanism by which these mutations contribute to CD susceptibility has yet to be fully elucidated. The location of the variants in the LRR region suggests that these mutants are defective in sensing microbial ligands and therefore in initiating activation of the NOD2 protein and NFκB signalling i.e. the mutations cause loss of function. This is supported by data from human monocytes harbouring NOD2

mutations which show decreased activation of downstream cytokine signalling (Negoro et al., 2005). Additionally, HEK293T cells transfected with NOD2 mutant constructs show decreased NFκB signalling (Inohara et al., 2003). Furthermore, macrophages from both NOD2 null and L1007fsInsC homozygous mice exhibit reduced NFκB activation following stimulation with MDP (Y.-G. Kim et al., 2011).

An alternative mechanism by which NOD2 mutations may influence CD susceptibility is via the autophagy pathway. NOD2 physically interacts with the autophagy protein ATG16L1 and recruits it to the plasma membrane. The NOD2 L1007fsInsC protein does not recruit ATG16L1 to the plasma membrane (Travassos et al., 2010). MDP stimulation of human cell lines shows increased bacterial killing upon *Salmonella* infection. NOD2 overexpression also results in increased bacterial killing in this model, furthermore, siRNA knockdown of *ATG16L1* blocks this increase. The R702W and G908R *NOD2* mutants exhibit decreased bacterial killing and bacterial killing is abolished in the L1007fsInsC mutant (Homer, Richmond, Rebert, Achkar, & McDonald, 2010). MDP has also been shown to induce autophagy in human dendritic cells expressing WT NOD2. *NOD2* knockdown in these cells results in decreased levels of the autophagy marker LCIII upon MDP stimulation. Furthermore, although CD associated *NOD2* mutations were not shown to alter autophagy in monocytic CD patient derived cells, they do alter autophagy in the dendritic cells of CD patients (Cooney et al., 2010; Homer et al., 2010). Dendritic cells cultured from CD patients with either ATG16L1 or NOD2 mutations have defects in bacterial clearance, autophagy and antigen presentation (Cooney et al., 2010).

NOD2 gain of function in Blau's Syndrome

In addition to LOF mutations in NOD2 being associated with Crohn's disease, gain of function (GOF) mutations have been associated with other auto immune diseases such as Blau Syndrome. Blau syndrome is an autosomal dominant disorder characterised by arthritis and dermatitis. Three missense mutations in the NOD domain of NOD2 are associated with Blau syndrome, these are thought to cause increased sensitivity of

NOD2 to ligands leading to increased activation of NFκB and other downstream signalling pathways (Miceli-Richard et al., 2001).

NOD2 null mice

The investigation of the functional significance of the CD associated *NOD2* mutations is complicated by *NOD2* null mice. These are mice lacking the NOD domain required for activation. These mice do not develop spontaneous inflammation and do not exhibit increased sensitivity towards dextran sodium sulphate (DSS) induced colitis. DSS is a chemical which destroys mucosal epithelial cells and disrupts the epithelial cell barrier therefore allowing bacterial infiltration of the mucosa (Kitajima, Takuma, & Morimoto, 1999). Additionally, there was no difference in bacterial clearance between wild type and NOD null mice when exposed to gram positive bacteria by intraperitoneal or intravenous injection only when exposed via oral routes (K. S. Kobayashi, Chamaillard, & Ogura, 2005). This phenotype is not necessarily surprising as there are numerous PRRs that function as part of the innate immune system, these may be compensating for NOD2 inactivation.

Paneth cell dysfunction and decreased expression of α defensins have been observed in NOD2 null mice (K. S. Kobayashi et al., 2005). Defensins are antimicrobial peptides, in the small intestine they are present in Paneth cells. This, in line with high levels of NOD2 expression in human Paneth cells, suggest that NOD2 plays a role in maintenance and protection of the intestinal mucosa which may be mediated by interaction with α defensins. Reduced expression of α defensins has been seen in ileal CD patients compared to healthy controls, however, more recently, this has been attributed to inflammation as opposed to NOD2 status (Simms et al., 2008; Wehkamp et al., 2005).

1.4.7 Yeast 2 Hybrid Assays

Linkage and GWA studies have identified many different susceptibility loci, the most significant of which is the *NOD2* locus. In the post GWAS era the research focus has shifted towards identifying the mechanisms underlying the associations discovered. In our lab one of the approaches was to identify proteins that interact with the known CD susceptibility gene: *NOD2*.

Yeast 2 hybrid assays have been used in other IBD studies to identify potential IBD risk genes. The DNA binding domain of a transcription factor (TF) is attached to a bait protein (usually the known protein of interest) and DNA activation domain of a TF is attached to prey proteins (usually encoded by a cDNA library), when the bait and prey interact, the TF DNA binding domain and activation domain form an active TF initiating transcription of a reporter gene (Figure 5). GRIM 19 was identified as an interacting partner of NOD2 using a yeast two hybrid assay, this interaction was confirmed by co immunoprecipitation using *GRIM19* and *NOD2* over expression constructs in COS7 and HEK293 cells (Barnich, Hisamatsu, et al., 2005). In this study the N terminal region of NOD2 lacking the CARD domain was used as bait and cDNA generated from human bone marrow as prey. ERBIN has also been shown to interact with NOD2 using a yeast two hybrid assay. In these experiments, full length *ERBIN* and *NOD2* were used and the interaction was confirmed by immunoprecipitation in SW480 cells.

TLE1 was not identified by these screens, this may be due to differences in cell lines and methodologies used. Nimmo et al used full length NOD2 extracted from SW480 cells as prey and co immunoprecipitation was confirmed in HEK293 cells.

This approach has also been used to understand pathogenic mechanisms and identify susceptibility genes in other diseases. Millar et al used a yeast 2 hybrid assay to successfully identify interacting partners of DISC1, which has been implicated in schizophrenia. They identified 21 interacting proteins, implicating DISC1 in a novel

pathways and identifying it as a potential hub protein (Millar, Christie, & Porteous, 2003). Interestingly, GWA studies for psychiatric illnesses have not identified *DISC1* as a susceptibility locus. However, GWA studies have identified interacting partners of DISC1, these are reviewed in (Brandon & Sawa, 2011).

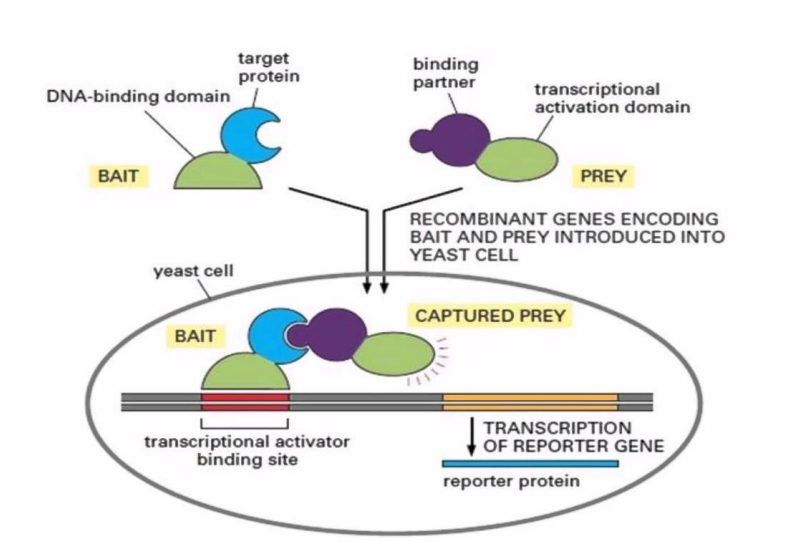


Figure 5 Yeast 2 Hybrid Assay

The bait protein is attached to the DNA binding domain of a transcription factor (TF). The prey which will be a protein generated from a large cDNA library is attached to a TF activation domain. If the bait and prey proteins interact, then the TF binding domain and TF activation domain interact in the yeast cells and function as a transcriptional activator to induce expression of a reporter gene.

1.5 TLE1 (Transducin- like enhancer of Split)

1.5.1 TLE1 in Crohn's Disease

Recent research by Nimmo et al identified TLE1 as an interacting partner of the known IBD susceptibility gene, NOD2 using a yeast-2 hybrid assay. This interaction was confirmed by co immunoprecipitation in the colorectal cancer cell line, SW480. Furthermore, three SNPs in *TLE1* were significantly associated with CD in a combined Scottish cohort (Edinburgh and Dundee): rs11139315 (p=0.003), rs2796469 (p=0.004) and rs6559629 ($p=4\times10^{-4}$). Statistical analysis on this combined cohort showed that the rs6559629 risk allele in *TLE1* was required for *NOD2* to be a risk factor for CD (Nimmo et al., 2011) . However, it is important to note that the latter has not been replicated in data from larger UK IBD cohorts (see chapter 6).

1.5.2 Discovery and structure of TLE1

TLE1 is a transcriptional co repressor and one of the human homologs of the Drosophila *groucho* family of proteins. *Groucho* is a nuclear protein, mutations in which cause the development of extra bristles above the eye, reminiscent of the eyebrows of Groucho Marx (Lindsley & Grell, 1968). *Groucho* was initially classified as part of the *enhancer of split complex* (*e(sp)l*) due to its location in the densely packed *e(sp)l* chromosomal region. This, combined with sequence similarity to the β subunit of the TRANSDUCIN protein lead to the name, *Transducin Like Enhancer of split* for the human *groucho*/*TLE1* family. Further work showed that functionally, *groucho* was not related to *β transducin* or the *e(sp)l* complex.

The TLE family consists of four proteins, TLE1-4 (Stifani, Blaumueller, Redhead, Hill, & Artavanis-Tsakonas, 1992). These proteins are highly conserved, both groucho and TLE proteins have the same five protein domains (Q, GP, CcN, SP and WD) as shown in Figure 6. The Q and WD domains are particularly well conserved between Groucho and TLEs, they share between 70% and 88% sequence identity. The WD and Q domains are where the majority of protein interactions take place. The coil structures in the Q domain are essential for oligomerisation of groucho/TLE, mutations in this region prevent oligomerisation and repression of target genes (Song, Hasson, Paroush,

& Courey, 2004). A number of proteins interact with the Q domain including the TCF/LEF transcriptional regulators (Chodaparambil et al., 2014). Along with the glycine/proline rich GP domain, the Q domain is also the site for binding histone deacetylases (Chen, Fernandez, Mische, & Courey, 1999; Choi, Kim, Kwon, & Kim, 1999; Palaparti, Baratz, & Stifani, 1997; Wang, Chen, & Ouyang, 2011).

The central CcN domain contains nuclear localisation signals and phosphorylation sites. The two serine residues in this domain are phosphorylated by cell cycle dependant kinase 2 (CDC2) and Casein Kinase 2 (CK2). Phosphorylation of serine residues in the CcN domain has also been shown to be important for efficient transcriptional repression and nuclear localisation of TLE1 (H. N. Nuthall, Joachim, & Stifani, 2004). Inhibition of *CDC2* causes dephosphorylation of TLE1 (H. Nuthall, Husain, & McLarren, 2002; H. N. Nuthall, Joachim, Palaparti, & Stifani, 2002).

The role of the SP domain of TLE proteins is also regulatory, cofactors such as HES1 induce phosphorylation of serine residues in this region. Mutations affecting serine 286 (SP domain) of TLE1 bind to chromatin with less strength that wild type TLE1, although TLE1 is still recruited to the chromatin (Buscarlet et al., 2009). HIP2K and MAPK have also been shown to induce hyperphosphorylation of serine residues in the SP domain, this phosphorylation is thought to alter the conformation of TLE1, exposing additional serine residues as well as negatively impacting transcriptional repression (Ciarapica et al., 2014; Hasson et al., 2005).

The WD domain is the site where a majority of protein interactions take place. This region has a β propeller structure with seven blades (WD repeat motifs) that has been resolved by X ray crystallography (Pickles, Roe, Hemingway, Stifani, & Pearl, 2002). The central pore region of the WD repeat region is where a majority of protein interactions occur (Pickles et al., 2002). The WRPW and Eh1 (Engrailed Homology region 1) motifs in this domain are essential to most of the TLE1 protein interactions. The WRPW motif is the recognition site for proteins from the basic helix loop helix

(bHLH) family (HES) and the Runt homology domain protein family (RUNX) (Ali et al., 2010; Grbavec & Stifani, 1996; McLarren, Theriault, & Stifani, 2001). The Eh1 site is where proteins including PAX and FOXD bind (Dastidar, Narayanan, Stifani, & D'Mello, 2012; Eberhard, Jime, Heavey, & Busslinger, 2000).

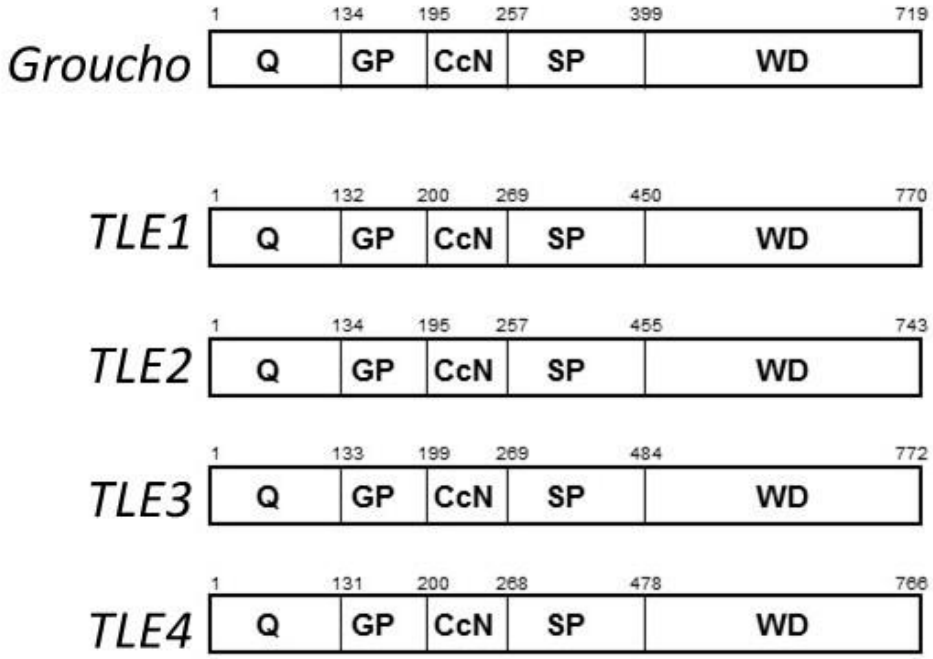


Figure 6 Structure of *Groucho* and full length TLE1 proteins

Protein domain structure of *Drosophila groucho* and human TLE proteins. Amino acid numbers are shown above each schematic diagram. The Q domain is the glutamine rich domain, the GP domain is glycine proline rich, the CcN is the central domain containing nuclear localisation signals, the SP is serine rich and the WD domain contains WD repeat sequences necessary for protein-protein interactions.

1.5.3 Expression patterns in different tissues

Human TLE (hTLE) proteins are ubiquitously expressed in human tissues. hTLE1 is highly expressed in the brain, liver and muscle, as determined by northern blot analysis (Stifani, Blaumueller, Redhead, 1992). TLE1 is primarily a nuclear protein, in line with its function as a transcriptional co repressor, however it has also been detected to a lesser extent in the cytoplasm in HeLa and HEK 293 cell lines (Nimmo et al., 2011; Stifani, Blaumueller, Redhead, 1992).

Overexpression of the TLE1 homolog, *Grg1*, causes defects in neuronal differentiation and pituitary gland development in mice. In Medaka embryos, overexpression leads to developmental defects including left/right asymmetry (reviewed in: (Buscarlet & Stifani, 2007))

1.5.4 TLE1 transcripts

TLE1 is located on chromosome nine, the full length transcript is 3893 base pairs in length and encodes a protein composed of 770 amino acids.

There are five TLE1 transcripts described by the Ensembl genome browser, these are shown in Figure 7. The Transcript Support Level (TSL) ranges from one to five, TSL1 indicates that all splice junctions are supported by RNA evidence, TSL2 means that the best supporting mRNA is suspect or support is from more than one expressed sequence tag (EST), TSL3 means there is only support from the transcript from one EST, TSL4 indicates the best supporting EST is suspect and TSL5 means no single transcript supports the model structure. Only transcript one has a TSL of 1, the other transcripts have TSL scores between 2 and 5 and have therefore not been properly validated yet.

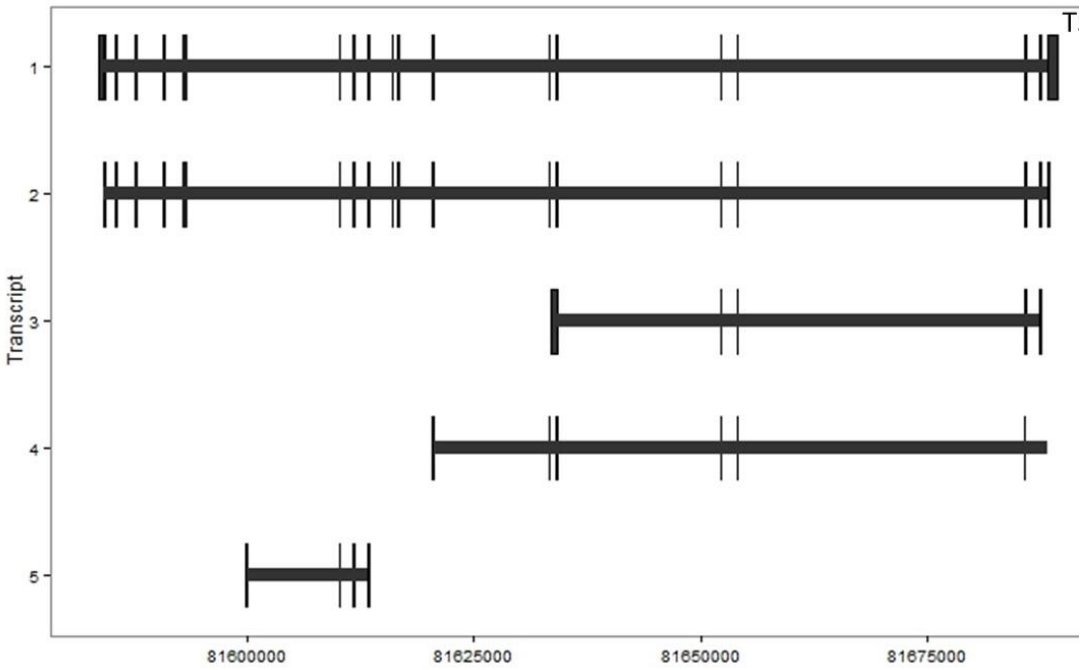


Figure 7 *TLE1* transcripts

TLE1 predicted transcripts, data from Ensembl genome browser. Transcript support level (TSL) shown for each transcript, definitions of each TSL level as described in text.

1.5.5 TLE1 function

TLE1 is a transcriptional co repressor, it lacks the ability to bind DNA directly but influences transcription by interacting with other DNA binding proteins. It is thought to be involved in long range repression, although the exact mechanisms underlying this repression are unclear. They are hypothesised to involve both chromatin remodelling and direct inhibition of transcription and translation (Courey & Jia, 2001). TLE1 appears to cause chromatin remodelling via interactions with the H3 subunit of histones, Histone deacetylase 1 (HDAC1) and Histone deacetylase 3 (HDAC3). The *Drosophila* histone deacetylase, *rpd3* interacts with *groucho* via the GP domain of the groucho protein. Cells overexpressing TLE1 recruit HDAC1 and knockdown of HDAC3 results in increased expression of *TLE1* intestinal epithelial cell lines (Mario F Fraga et al., 2008a; Godman et al., 2008). TLE1 has been shown to associate with H3 in human cell lines, this interaction is thought to occur at the amino terminus of H3

(Palaparti et al., 1997). *TLE1* inhibits protein translation via interaction with the RUNX2 proteins, RUNX2 and TLE1 form a complex which represses transcription of rRNA genes. It has also been shown that loss of TLE1 leads to increased global protein synthesis (Ali et al., 2010).

1.5.6 TLE1 in different signaling pathways

TLE1 functions as a transcriptional co-repressor in a variety of different cellular and developmental pathways e.g. NFκB, Wnt, TGFβ and Notch signalling pathways (Fisher, Ohsako, & Caudy, 1996; Levanon et al., 1998; McLarren et al., 2001). TLE proteins cannot bind DNA directly, they influence transcription via interaction with DNA bound transcription factors and the recruitment of histone deacetylases e.g. Sirt-1 (Ghosh, Spencer, Ng, McBurney, & Robbins, 2007a).

TLE1 in the Transforming growth factor β (TGFβ) signalling pathway

Maintaining a balance between pro-inflammatory and anti-inflammatory factors is one of the essential functions of the immune system, disturbance of this balance leads to inflammation. As previously mentioned, chronic inflammation is one of the major characteristics of IBD. In the GI tract the maintenance of this balance is dependent on effective communication between intestinal epithelial cells and the mucosal immune system. One of the ways in which this is achieved is via signalling molecules such as cytokines. Pro-inflammatory cytokines, for example, TNFα are important for the maintenance of inflammation in IBD patients, as shown by the use antibodies against these cytokines as treatment for IBD e.g. Infliximab. TNFα is a pro inflammatory cytokine that is overexpressed in the mucosa of IBD patients. It is primarily produced by macrophages, monocytes and T cells and TNFα receptors are broadly expressed. In response to TNFα stimulation, the MAP kinase and NFκB signalling pathways are initiated and inflammatory cells are recruited. In IBD patients, Infliximab (TNFα antibody) is used to sequester TNFα and reduce inflammation. Infliximab is also used to treat other autoimmune diseases including rheumatoid arthritis and ankylosing spondylitis.

TGF β is a regulatory cytokine and has a variety of functions which include suppressing local immune responses to antigens in the lumen, enhancing barrier function and increasing production of immunoglobulins in the *mucosa*. TGF β binds to TGFβ cell surface receptors, leading to activation of the receptor complex. This in turn causes downstream phosphorylation and activation of TGF β – associated signalling molecules (SMAD's)**.** Activated SMAD complexes then translocate into the nucleus and activate or repress specific target genes (Figure 8). One of these target genes is *Brinker*; this gene is down regulated by the SMAD complex. Brinker been shown to interact with the *Drosophila* TLE1 homolog, Groucho.

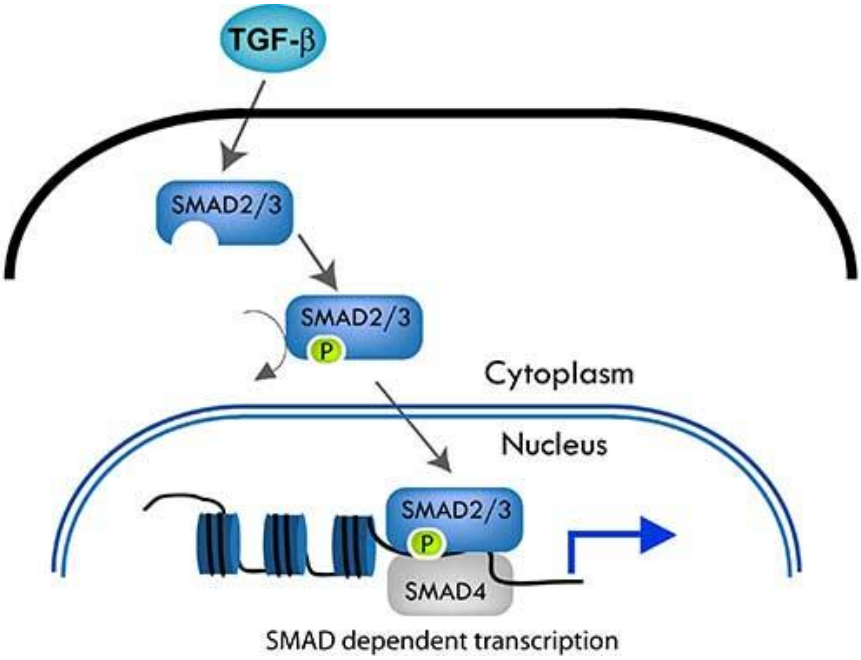


Figure 8 TGFβ signalling pathway

TGFβ binds to cell surface receptor molecules which leads to phosphorylation and subsequent activation of downstream SMAD molecules. There are three types of SMAD's: receptor regulated (R SMAD's 2, 3), co-mediator (Co SMAD's 4) and inhibitory SMAD's (SMAD 7). Upon binding of the TGFβ ligand to cell membrane receptors, an activated SMAD complex, consisting of an R SMAD and Co SMAD is formed. This complex translocates into the nucleus and interacts with DNA and other regulatory proteins to influence transcription.

Analysis of colonic biopsies from healthy controls has suggested that TGFβ is important in down regulation of inflammation, exposing these biopsies to TGFβ antibodies increased production of pro-inflammatory cytokines e.g. TNFα (Reimund, Wittersheim, Dumont, & Muller, 1996). However, colonic biopsies from IBD patients have shown increased TGF β expression. Recent research has shown that this is due to over expression of SMAD7, this inhibitory SMAD binds to the TGF β receptor which in turn stops activation of SMAD 3 and the formation of an activated SMAD complex. Inhibition of SMAD3 has been shown to restore normal TGF β signalling.

TLE1 in the Notch signalling pathway

The Notch signalling pathway is an evolutionarily conserved pathway that controls cell fate determination during development e.g. haematopoiesis, somitogenesis, vasculogenesis and neurogenesis. These processes involve maintenance of stem cell self-renewal, proliferation, specification of cell fate or differentiation and apoptosis. Mutations that alter Notch signalling have been implicated in a number of developmental disorders and cancers including colorectal cancer (Qiao & Wong, 2009).

Notch signalling is initiated by binding of ligands e.g. DELTA and JAGGED to the NOTCH transmembrane receptor. Upon NOTCH activation γ secretase cleaves the Notch intracellular domain (NICD). The NICD translocates into the nucleus where it interacts with the CSL protein and other regulatory proteins to activate transcription of target genes including the *HES* genes (Figure 8). Research in *Drosophila melanogaster* has shown that *groucho* is a *notch* antagonist; when notch is inactive, *groucho* inhibits transcription of the csl proteins (Nagel et al., 2005). TLE1 has been shown to interact with HES proteins to mediate transcriptional repression (Grbavec & Stifani, 1996).

Maintenance and renewal of the intestinal epithelium is governed by intestinal stem cells, *Notch* is highly expressed in these cells (Okamoto et al., 2009). These cells are

situated at the base of the intestinal crypt, they progressively differentiate, migrate and proliferate along the crypt villus axis, eventually becoming functional absorptive, goblet or epithelial cells of the villus. In inflamed mucosa of IBD patients this process is often dysregulated and an excess of dedifferentiated or differentiated cells can be seen (Dahan et al., 2008). In mice, notch is essential for the proliferation and differentiation of intestinal epithelial cells it and inhibition of notch causes increased severity of colitis (Okamoto et al., 2009). NOTCH binds to TNFα and is a target of anti TNFα therapies such as Infliximab that are used to treat IBD. Additionally, *Notch* is activated and upregulated in patients treated with anti- TNFα drugs (Werner et al., 2012).

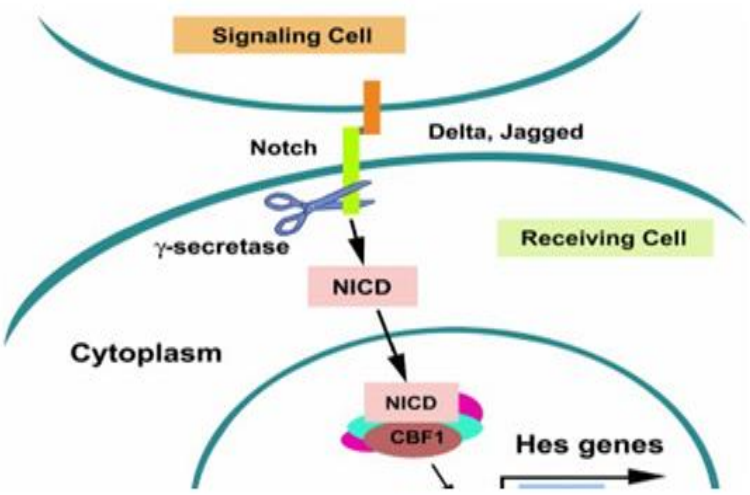


Figure 9 Notch signalling pathway

Notch is a transmembrane receptor protein that is activated by ligands of the Serrate/Lag2 family which include the DELTA and JAGGED proteins. Ligand binding results in cleavage of the Notch intracellular domain (NICD). The NICD translocates from the cytoplasm into the nucleus where it interacts with CBF/Su(H)/Lag2 (CSL) proteins. The NICD and CSL along with other co-activators form a complex which binds DNA and functions as a transcriptional activator. Targets of this complex include the *HES* genes.

Recent research has further implicated the Notch pathway in IBD pathogenesis, the activated form of NOTCH has been shown to be over expressed in the epithelium of CD patients when compared to controls (Dahan et al., 2008). CD patients have fewer

goblet cells than healthy controls (Gersemann et al., 2009). Furthermore, over expression of the NICD in goblet cell lines causes an abnormal goblet cell phenotype to develop, further confirming the importance of the Notch pathway in IBD (X. Zheng et al., 2011).

TLE1 in the Wnt signalling pathway

The Wnt–β-catenin signalling pathway plays a critical role in development, cell fate determination and adult stem cell proliferation. The abnormal activation of the Wnt–β-catenin signalling pathway has been found to be associated with a number of human cancers, including colorectal cancer (Jin et al., 2003; Korinek, 1997; Krings et al., 2000; Morin, 1997; Polakis, 2012; Rubinfeld, 1997).

The Wnt signalling pathway is activated by the presence of Wnt glycoproteins which bind to and activate FRIZZLED, and low density lipoprotein receptors (LDL 5 and 6). This causes accumulation of β CATENIN in the nucleus which forms a complex with the TCF/LEF proteins and other co-activators to initiate transcription activation (Figure 10). In the absence of Wnt signalling the TCF/ LEF proteins are bound by TLE1 and Histone deacetylase 1 (HDAC 1) in a transcriptional repression complex (Daniels & Weis, 2005).

The Wnt signalling pathway has been found to be dysregulated in UC. The *mucosa* of UC patients has shown over expression of a number of Wnt pathway related genes including the Wnt receptor *FRIZZLED* (Uthoff et al., 2001). Additionally both epigenetic and genetic alterations in this pathway have been shown to be involved in the development in colorectal cancer in UC patients (Dekken et al., 2007).

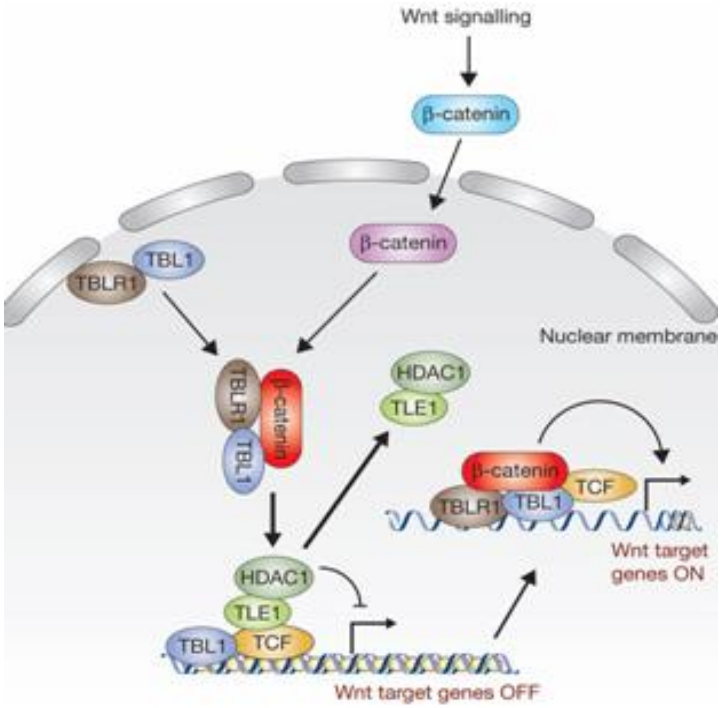


Figure 10 Wnt signalling pathway

Wnt's are glycoproteins that interact with the cell surface receptors FRIZZLED and low density lipoprotein receptors 5 and 6 (LDL 5 and 6). Activation of these Wnt receptors leads to accumulation of the transcriptional co activator β CATENIN in the nucleus. In the absence of Wnt signalling the TCF/LEF proteins are bound by histone deacetylase 1 (HDAC1) and TLE1. In the presence of Wnt signalling, TLE1 and HDAC1 are displaced by β CANTENIN and a transcriptional activation complex is formed.

TLE1 in the NFκB signalling pathway

TLE1 proteins have been implicated in the NFκB pathway, TLE1 directly interacts with and represses transcription activity of the p65 subunit of NFκB dimers (Tetsuka et al., 2000). Additionally *Groucho* interacts with *Dorsal*, the *Drosophila* NFκB homolog, converting *Dorsal* into a transcriptional repressor (Dubnicoff et al., 1997). A yeast 2 hybrid screen identified TLE1 as an interacting partner of the deacetylase and NFκB inhibitor, SIRT1. This finding was confirmed by co-immunoprecipitation of TLE1 and SIRT1 in HeLa cells. Furthermore, although SIRT1 does not directly influence expression of TLE1 or vice versa, TLE1 is necessary for NFκB inhibition by SIRT1 (Ghosh, Spencer, Ng, McBurney, & Robbins, 2007b).

Inflammation is a hallmark of IBD, it is seen in affected regions of the GI tract in both CD and UC patients. As the NFκB pathway is a pro-inflammatory pathway it has been well studied in IBD and is a therapeutic target for IBD treatments. NFκB (p65) is overexpressed in the inflamed intestines (macrophages and epithelial cells) of IBD patients and the extent of NFκB (p65) expression correlates with the extent of inflammation (Rogler et al., 1998). The known IBD susceptibility gene *NOD2* also induces NFκB activation through CARD domain interactions with RICK, as discussed earlier in this chapter (Yasunori Ogura et al., 2001).

A number of current IBD treatments including anti TNFα antibodies (Infliximab) and corticosteroids are thought to alleviate IBD symptoms by targeting the NFκB pathway. The anti-inflammatory effects of anti-TNFα treatments are at least partially attributable to inhibition of the NFκB p65 subunit and increased expression of catalytic components of the IKK complex (Guidi et al., 2005). Corticosteroids are also thought to increase expression of the catalytic components of the IKK complex, therefore increasing inhibition of NFκB proteins and preventing translocation into the nucleus, they have also been shown to directly interact with the NFκB p65 subunit (Auphan, Nathalie, Didonato, Joseph, Caridad, Rosette, Helmburgh, 1995).

1.6 Dysregulation in cancer

TLE1 expression has been studied in a range of cancers including: breast, bone/joints (sarcomas), lung, colon and pancreatic cancer (Allen et al., 2006; Hamidov et al., 2011; Knösel et al., 2012; Kosemehmetoglu, Vrana, & Folpe, 2009). In hematologic malignancies TLE1 has been shown to undergo promoter hypermethylation (Mario F Fraga et al., 2008b).

UC and CD predispose to colorectal cancer (CRC) and lead to a poorer prognosis (reviewed in: (Kim & Chang, 2014)). In approximately half of the CRC cases histone

deacetylase 3 (HDAC3) is overexpressed. In the CRC cell lines, it has been shown that inhibition of HDAC3 causes increased expression of TLE1 and other members of the Wnt signalling pathway (Godman et al., 2008).

1.6.1 TLE1 deletion in model organisms

Groucho null *Drosophila melanogaster* develop extra bristles above the eyes and their eyes a show developmental abnormalities. *Groucho* mutants also show abnormalities in neurogenesis and sex determination (Paroush et al., 1994). *Gro* null mice are born healthy, however they show growth abnormalities from three days old onwards with a survival of 50% by the age of four weeks when compared to healthy control littermates. In response to TLR ligand stimulation, macrophages derived from these mice show increased levels of TNFα and IL6 when compared to macrophages derived from healthy littermates (Ramasamy, Chen, Wang, Ding, & Sweetser, 2010).

1.7 Aims

As discussed in this chapter, *TLE1* has been implicated in CD pathogenesis by way of its interaction with NOD2 and association with CD in a case control GWA study (Nimmo et al., 2011). Additionally *TLE1* functions in signalling pathways that *NOD2* has been implicated in such as the NFκB pathway. Furthermore *TLE1* functions as part of the Wnt, Notch and TGFβ pathways, all of which have previously been implicated in IBD. The aim of this thesis is to elucidate the mechanism by which *TLE1* may be influencing susceptibility to CD. The work presented analyses *TLE1* sequence, expression and interacting proteins in order to further understand the function of *TLE1* and its potential role in CD pathogenesis. This in turn may allow for identification novel therapeutic targets leading to improved quality of life for CD patients and lessening the current burden on the NHS.

The specific aims of each results chapter of this thesis are as follows:

1. To analyse the effect of down regulating *TLE1* expression on genome wide expression to identify mechanisms by which *TLE1* may be influencing susceptibility to CD.

2. To examine the potential *TLE1*/ XBP1 interaction and its role in the unfolded protein response (UPR).

3. To determine TLE1 expression and localisation in ileal tissue taken from healthy controls and CD cases.

4. To sequence regions of the *TLE1* gene to identify causative mutations underlying the association between *TLE1* and CD in the Scottish population.

2 Materials and Methods

2.1 Buffers and solutions

2.1.1 Bacterial culture

L Agar	50g Tryptone 25g Yeast extract 50g NaCl 5L dH_2, pH 7.2
L Broth	50g Tryptone 25g Yeast extract 25g NaCl 5L dH_20, pH 7.2
Glycerol freezing medium	100ml L Broth 25% glycerol

Table 1 Solutions required for bacterial culture

2.1.2 Cell culture

Phosphate Buffered Saline (PBS)	Tablets (Invitrogen) dissolved in 1L dH_2O pH 7.5 Tablet contains: 10mM Phosphate 150mM NaCl
DMSO freezing medium	10% Dimethylsulfoxide (DMSO) 90% Fetal calf serum (FCS)

Table 2 Cell culture buffers and solutions

Blocking Buffer	0.1% Tween (Sigma) 5% w/v powdered milk (Marvel) 10ml PBS

PBS- Tween (PBS-T)	PBS supplemented with 0.1% tween

Table 3 Buffers and solutions for protein analysis

2.1.3 Chromatin Immunoprecipitation

ChIP Dilution Buffer	16.7mM TrisHCL pH8.1 (3.3ml, 0.5M) 1.2mM Ethylinediamenetetraacetic acid (EDTA) (0.34ml, 0.5M) 167mM NaCl (4.15ml,4M) 1.1% Triton 0.01% Sodium dodecyl sulphate (SDS) (0.1ml, 10%) 91.1 ml H_2O
ChIP Lysis Buffer	50mM TrisHCl pH8.1 (10ml, 0.5M) 10mM EDTA (2ml, 0.5M) 1% SDS (10ml, 10%) 78ml distilled H_2O
ChIP Elution Buffer	0.1M $NaHCO_3$ (0.084g) 1% SDS (1ml, 10%) 9ml H_2O
TSEI	20mM TrisHCl pH8.1 (4ml, 0.5M) 2mM EDTA (0.4ml, 0.5M) 150mM NaCl (3.8ml, 4M) 1% Triton (1ml) 0.1% SDS (1ml, 10%) 89.9 ml H_2O
TSE II	20mM TrisHCl pH8.1 (4ml, 0.5M) 2mM EDTA (0.4ml, 0.5M) 500mM NaCl (12.5ml, 4M) 1% Triton (1ml)

	0.1% SDS (1ml, 10%) 81.3ml H_2O
TE	50mM Tris HCL pH8.1 (10ml, 0.5M)
Buffer III	1mM TrisHCl pH8.1 (2ml, 0.5M) 1mM EDTA (0.2ml, 0.5M) 1% NP40 (5ml, 5%) 1% 2-4 dimethyoxy-4 –chloramphetamine (DOC) (10ml, 10%) dH_2O (7.8ml)

Table 4 Buffers used in chromatin immunoprecipitation experiments (ChIP)

2.1.4 Antibodies

Antibody	Description	Species	Concentration	Application
M101 (Santa Cruz)	Anti TLE1	Rabbit	1:20	Immunohistochemistry (IHC)
M186 (Santa Cruz)	Anti XBP1 (spliced and unspliced)	Mouse	1:100	Western blotting and Chromatin Immunoprecipitation (ChIP)

Table 5 Primary antibodies used in this thesis

Name/Target	Concentration	Application
HRP donkey anti mouse (DAKO)	1:1000 1:100	Western Blot ChIP

Table 6 Secondary antibody used in this thesis

2.1.5 Expression vectors

	Vector	Promoter	Resistance	Source
Empty vector		CMV	Ampicillin	(Nimmo et al., 2011)
TLE1		CMV	Ampicillin	(Nimmo et al., 2011)
NOD2		CMV	Ampicillin	(Nimmo et al., 2011)
XBP1(unspliced)		CMV	Ampicillin	(Kaser et al., 2008)
XBP1 (spliced)		CMV	Ampicillin	(Kaser et al., 2008)

Table 7 Summary of expression constructs used in this thesis

2.2 Bacterial Culture

Plasmids were transformed into competent *E.Coli* cells (Life Technologies) for amplification. Cells were transformed with 1ul of plasmid. Briefly, 50µl of cells were mixed with 1µl plasmid DNA (approximately 20ng), the mixture was incubated in ice for 30 mins, heat shocked at 42˚C for 30 seconds, 1ml LB broth was added and mixture was incubated in a shaking incubator for 90 minutes (37˚C, 369G) 100µl of this mixture was then spread onto agar plates containing the relevant antibiotic. The following day a single colony was picked and added to 250ml LB broth supplemented with the relevant antibiotic and incubated in a shaking incubator overnight (37˚C, 369G). Plasmid DNA was extracted using a Qiagen midiprep kit according to standard protocols. Plasmid concentration was analysed using a Nanodrop 1000 spectrophotometer (Thermo Fisher). Plasmid DNA was stored at -20˚C. Glycerol stores were made for long term storage of each plasmid by spinning down 1 ml of single colony bacterial culture, resuspending in 1ml glycerol freezing medium and storing at -80˚C.

2.3 Cell culture

2.3.1 Maintenance of cell lines

Cells were kept in a Galaxy 170S incubator at 37°C in a humidified environment of 5% CO_2 and 95% air. HEK293, HCT116, HT29 and SW480 cells were originally purchased from the American Type Culture Collection (ATCC). Cells were kept in 75cm or 125cm flasks and passed every 3-5 days (80% confluency). Unless otherwise stated, all cell treatments including transfection, stimulation and passing was conducted in a sterile hood with laminar air flow. Passing cells involved aspiration of media, washing in 10ml Phosphate Buffered Saline (PBS) and 5 minute TrypLE Express treatment (37˚C/ 5% CO_2) (Life Technologies). Media and PBS were kept at 4°C but were heated to 37°C in a water bath prior to use. Following treatment with either 2ml (25cm flask) or 5ml (75cm flask) trypsin either 10 or 20ml of the appropriate media was added to the flask to quench the trypsin. Cells were passaged at dilutions specified in Table 2. Cells were cultured in media purchased from life technologies according to Table 8. All glassware was autoclaved prior to use, all plastic ware was purchased pre-sterilised. Cells were tested for mycoplasma once a month, no cells were positive. Additionally cell morphology was analysed every week using a light microscope (Leica).

2.3.2 Cell counting

Prior to an experiment, once cells had been re suspended in media following TrypLE express treatment they were counted using a Cellometer X1 (Nexcelom). According to standard protocols, 20µl of cell solution was pipetted into each well of the counting chamber, the slide was placed in the cellometer, the image focused and the cell type chosen and counted. The software returned cell images which were analysed for counting accuracy and the number of cells per ml of cell solution.

Cell Line	Description	Dilution factor (passaging)	Media
HEK 293	Human Embryonic Kidney	1:20	DMEM + 10% FCS
HCT 116	Colorectal carcinoma	1:10	Mc Coy's + 10% FCS
HT 29	Colorectal carcinoma	1:10	DMEM + 10% FCS
SW 480	Colorectal carcinoma	1:10	Leibovitz +10% FCS

Table 8 Cell lines, Culture medium and passing dilution factor

2.3.3 Long term storage of cells

For long term storage, a T75 flask of cells was trypsinised and resuspended in the relevant media. The cell solution was precipitated (4°C, 1000RPM) and cells were resuspended in 1ml DMSO freezing medium and stored in cryovials. Vials were placed in a Mr Frosty containing isopropanol (Thermo Fisher) and kept at -80°C for 2-3 days. Vials were then transferred to liquid nitrogen stores at the Human Genetics Unit (HGU).

Cell stocks were thawed for use in a 37°C water bath for 3 minutes. Cell solution was then transferred to a T75 flask containing 10ml normal cell media. The following day, media was replaced to remove DMSO containing media.

2.3.4 Transient transfection

Cells were plated 24 hours prior to transfection in 10cm plates at a confluencies of: 2×10^6 (24 hour transfection), 1×10^6 (48 hour transfection) 5×10^5 (72 hour transfection). All transfections were carried out in 10cm plates. Lipofectamine 369G0 (Life Technologies) was used as a transfection reagent. 10µg of DNA construct was mixed with 500µl of Opti-MEM (Life Technologies), the solution was incubated for 5 mins (room temperature (RT), sterile hood). Simultaneously, 18µl of Lipofectamine was added to 500µl of Opti-MEM and incubated for 5 minutes (RT, sterile hood). The DNA solution was then mixed with the Lipofectamine solution and incubated for 20 minutes (RT, sterile hood). During this the media of cells to be transfected was removed and replaced with 9ml of the appropriate fresh media. Following 20 minute incubation, the DNA-Lipofectamine mix was added dropwise to the 10cm plate of cells.

After the relevant time period for transfection had passed, cells were washed in PBS at the bench, scraped in 2ml PBS and pelleted by centrifugation 9167G). PBS was removed and pellets were stored in Eppendorf tubes at -20°C prior to RNA, DNA or protein extraction.

2.3.5 MDP stimulation

Cells were plated in 10cm plates according to standard protocols 24 hours prior to stimulation. The following day, media was aspirated from the cells and replaced with 10ml of the relevant media supplemented with 10µl L18 MDP (Invitrogen). Following 6 or 24 hour incubation in a humidified incubator (37˚C, 5% CO_2) cells were washed in PBS. Cells were scraped into 2ml PBS using a cell scraper and pelleted by centrifugation (4˚C, 10,000 RPM).

2.4 Protein analysis

2.4.1 Sample preparation

Cell pellets were lysed by addition of 200µl Lamelli loading buffer (Biorad). The solution was sonicated for at a 10 micron amplitude for 1 minute on ice and then heat treated at 45°C for 10 minutes. Finally, samples were vortexed for 30 seconds, ready for loading.

2.4.2 Western Blotting

Electrophoresis

20µl of cell lysate was loaded onto a 10% NuPage SDS-PAGE gel (Invitrogen) alongside PageRuler plus protein ladder (Thermo Fisher). The western blotting tank gel chamber was filled to the brim and the tank was half filled with NuPage buffer (Invitrogen). This was run at 150V (mA set to maximum) for 10 minutes and then 120V for approximately 45 minutes, until the 25KDa ladder was at the end of the gel.

Transfer

Following electrophoretic separation of proteins on the gel, proteins were transferred to a nitrocellulose membrane (GE Healthcare). This was performed using the Mini Trans Blot system (Biorad). Briefly, nitrocellulose membrane, two sponges and two pieces of filter paper were soaked in 1X transfer buffer (Invitrogen). A sponge was placed on one side of the plastic cassette, followed by a piece of filter paper. The gel was extracted from its plastic cassette and placed on the filter paper. The nitrocellulose membrane was placed on top of the gel, then an additional piece of filter paper and the final sponge. Any bubbles in the "sandwich" were removed by rolling a 10ml plastic stipette over it. The cassette was then placed in the transfer apparatus. An ice block was placed in the tank alongside the transfer cassette. The entire transfer tank was filled to the brim with transfer buffer. Gels were transferred for 90 mins (RT). Protein transfer was confirmed by staining the nitrocellulose membrane with Ponceau S.

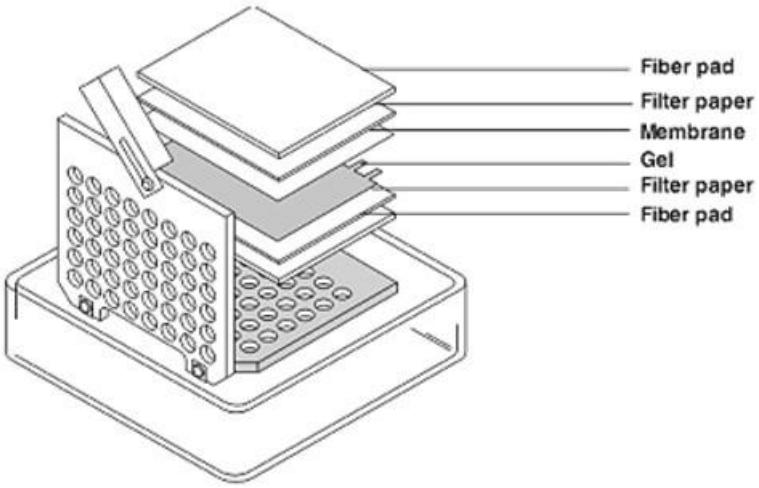


Figure 11 Western blot transfer

Adapted from Biorad. Schematic diagram of western blot transfer "sandwich". Assembled using filter paper, sponges, nitrocellulose membrane and polyacrylamide gel.

Immunoblotting

Ponceau S stain was removed by washing with PBS. Membranes were incubated in blocking buffer for 1 hour (RT). Membranes were then incubated in 10ml PBS supplemented with 5% milk powder and primary antibody, overnight at the

concentrations specified. The following day the membrane was washed in PBS-tween (PBS-T) for 5 mins, three times (RT). HRP conjugated secondary antibody was added at a concentration of 1:1000 (DAKO) for 1 hour (RT). Membranes were washed in PBS-T for 5 minutes, three times (RT). This was followed by chemiluminescent detection of proteins using an ECL kit (GBI Labs) with standard protocols. The membrane was wrapped in cling film and placed in a light poof cassette and was exposed to X-ray film for periods of 30seconds, 1 minute and 5 minutes. X-ray film was developed using a Curix 60 film processor (AGFA Healthcare).

2.5 RNA analysis

2.5.1 RNA extraction and cDNA conversion

RNA was extracted on the bench using an All Prep RNA/DNA kit (Qiagen) with standard protocols. At the final stage RNA was eluted in 30µl DEPC treated water. Extraction was performed on the bench, samples were kept on ice at all times. RNA was quantified using the Nanodrop spectrophotometer 1000 (Thermo Fisher) and RNA was converted to cDNA using a Superscript cDNA VILO kit (Invitrogen). RNAse free Eppendorf tubes were used at all stages of the conversion. Briefly, 1µg RNA was mixed with 2µl Superscript enzyme and 4µl 5X VILO reaction mix, DEPC treated water was added to 20µl. The mixture was incubated at 25˚C (10 mins) and 42˚C (60 mins). The reaction was terminated by incubation at 85˚C (5mins).

2.5.2 Real-time Quantitative PCR (RT qPCR)

RT qPCR was performed using a Rotor gene 6000 (Corbett Life Sciences/Qiagen). All water used in dilutions and RT qPCR negative controls was obtained from a Milli-Q Integral dispenser (Merck Millipore). Taqman expression assays were used to analyse *NOD2* (hs00223394_m1), *TLE1* (hs00896130_g1) and *GAPDH* (hs02758991_g1) expression (Thermo Fisher). Serial dilutions of cDNA (1:10-1:1000) were used in these assays. Briefly, 4µl diluted cDNA or ultrapure water (negative control) was added to a mixture containing: 1µl 20X Taqman Gene Expression Assay, 10µl 2X

Taqman gene expression master mix and 4µl ultrapure water. cDNA was added to the mixture in RT qPCR tubes and pipetted up and down. Taqman reagents and cDNA were kept on ice throughout the experiment. *RIOK1*, *SGPL1*, *TUSC3*, *CCND1*, *TBP*, *GAPDH* (not used when analysing *TLE1* or *NOD2* expression), *UBC*, *β ACTIN* expression and the XBP1 ChIP DNA enrichment RT qPCRs were analysed using a SYBR green based assay (DyNAmo Flash, Thermo Fisher). In the SYBR green assays reaction mixtures consisted of 10µl 2X master mix, 1µl 10µM of each primer (forward and reverse), 2µl of cDNA dilution/undiluted DNA (ChIP DNA enrichment) and ultrapure water to 20µl. RT qPCR reaction mixtures were prepared on a cooled rack. For each gene expression assay a standard curve was also run using serial cDNA dilutions ranging from 1-1:10000. The same standard curve samples were used for all runs of an experiment. RT qPCR reactions were placed in the Rotor gene 6000 and run on programs detailed in Table 9 and Table 10.

	Temp	Time
Hold	95˚C	7 mins
45 cycles	95˚C 56˚C	10 30
Melt	60-98˚C	90 secs per 1˚C step 5 secs between each step

Table 9 Gene expression RT qPCR program

	Temp	Time
Hold	95˚C	7 mins
45 cycles	95˚C 60˚C	10 30
Melt	60-98˚C	90 secs per 1˚C step 5 secs between each step

Table 10 ChIP RT qPCR program

RT qPCR data was analysed using Rotor gene analysis software (Corbett Life Sciences/Qiagen). A standard curve was used to generate concentration values for each sample. Concentrations were standardised to the relevant controls i.e. samples overexpressing *TLE1* were standardised to empty vector transfections. If negative controls showed contamination results for the run were discarded and re run.

2.6 DNA Analysis

2.6.1 DNA Extraction

DNA was extracted from cell lines using an All Prep DNA/RNA kit (Qiagen) with standard protocols. DNA pellets were resuspended in 30µl TE and was quantified using a Nanodrop spectrophotometer 1000 (Thermo Fisher).

DNA patient samples had already been extracted using the blood DNA Nucleon extraction kit (Hologic)

2.6.2 Polymerase Chain Reaction (PCR)

DNA samples to be used in PCR reactions were kept at 4˚C. DNA was used at a concentration of 100ng/µl. PCR reagents were defrosted and kept on ice during the experiment. A Hot start Taq polymerase kit (New England Biosciences) was used to set up PCR reactions. The reaction mixture contained: 2.5µl Taq reaction buffer,

0.5µl 10mM dNTPs, 0.5µl 10µM forward primer, 0.5µl 10µM reverse primer, 0.125µl Hot start Taq polymerase, 1µl 100ng/µl DNA and DEPC treated water to 25µl. When more than one PCR reaction was being performed, a master mix without primers or DNA was made up and pipetted into the PCR plate containing the relevant primers and DNA in each well. Thermocycling conditions are shown in Table 11.

	Temperature	Time
Initial denaturation	95 ˚C	30
30 cycles	95 ˚C 50-65 ˚C 68 ˚C	30 60 1 min/kB
Final extension	68 ˚C	5 mins
Hold	4 ˚C	

Table 11 PCR program specifications

2.6.3 Sanger sequencing

PCR products were sent for PCR cleanup and sequencing at the Human Genetics unit. 10µM primer stocks were provided with each sample. Primers sequences are shown in Appendix 1. Sequence chromatograms were analysed using Geneious 7.1.4 software.

2.7 Statistical Analysis

All statistical analysis was conducted using Microsoft Excel or R. All error bars represent the standard error of the mean, unless otherwise stated. Data from samples sizes less than 5 was assumed to have non- normal therefore non- parametric statistical tests (Mann U Whitney) were used to analyse these datasets.

3 Analysing the effect of knocking down *TLE1* and stimulating *NOD2* on genome wide expression

3.1 Introduction

3.1.1 Identification of direct and indirect *TLE1* targets

As previously discussed, *TLE1* is a transcriptional corepressor that has been implicated in IBD pathogenesis (Nimmo et al., 2011) . Transcriptional cofactors such as TLE1 influence transcription of large numbers of target genes ; *TLE1* functions as part of numerous developmental signalling pathways: NFκB, Wnt and TGFβ (Fisher et al., 1996; Levanon et al., 1998; McLarren et al., 2001).

In the first part of this chapter, *TLE1* expression was altered in a human cell line and the effect of this on genome wide expression were analysed using an Illumina HT12 expression chip. Analysis of the effect of *TLE1* knockdown will allow for identification of novel direct and indirect targets as well as novel pathways that *TLE1* may be involved in. This in turn may help to elucidate the mechanisms and signalling pathways by which *TLE1* may be influencing susceptibility to CD.

3.1.2 Functional characterisation of the relationship between *TLE1* and *NOD2*

MDP and *NOD2* stimulation

NOD2 is a pattern recognition receptor (PRR). When activated, it initiates downstream signalling pathways including the NFκB signalling pathway (as discussed in chapter 1). Muramyl dipeptide (MDP) is a peptidoglycan, which is a component of the bacterial cell wall of both gram positive and negative bacteria. MDP has been shown to bind to NOD2 and stimulate *NOD2* expression (Marian C Aldhous et al., 2011; Girardin et al., 2003; Grimes et al., 2012; Rosenstiel et al., 2003). The *NOD2* promoter contains NFκB regulatory sites and stimulation of cells with MDP initiates a positive feedback loop (Rosenstiel et al., 2003).

Giradin et al showed that MDP initiated *NOD2* dependant NFκB signalling. This was shown using a luciferase assay. In brief, this assay involved cloning the NFκB promoter region into a luciferase reporter plasmid. This plasmid was then transfected into HEK293T cells. Upon binding of RNA polymerase to the NFκB promoter,

luciferase is transcribed. Binding is quantified by lysing the cells, adding the luciferase substrate, luciferin and quantifying chemiluminescence. A 14 fold increase in NFκB activation was observed in cells transfected with *NOD2* and stimulated with MDP versus unstimulated cells transfected with *NOD2* (Girardin et al., 2003). MDP has been used successfully as *NOD2* stimulant in previous work conducted in our lab. A two fold increase in *NOD2* expression in cell lines stimulated with MDP for 4 hours was observed using RT qPCR (Marian C Aldhous et al., 2011).

As previously discussed, TLE1 was shown to interact with NOD2 using a Y2H assay. The pathological significance of this interaction has yet to be fully elucidated (Nimmo et al., 2011). As MDP is a component of bacterial cell walls, MDP was used as a *NOD2* stimulant in order to mimic the response of the host organism to bacterial exposure.

In the second part of this chapter MDP was used as a *NOD2* stimulant in human cells to determine whether increasing *NOD2* expression directly alters *TLE1* expression and then to analyse the effect of stimulating *NOD2* expression on genome wide expression. MDP stimulation was also combined with *TLE1* knockdown to identify any novel pathways that both *NOD2* and *TLE1* influence and to see whether any of the effects of *NOD2* stimulation rescue *TLE1*. These experiments may suggest a mechanism by which the NOD2/TLE1 interaction may be influencing susceptibility to CD.

3.1.3 Illumina HT12 Expression chip

The Illumina HT12 expression chip (beadchip) is an established method of analysing genome wide expression. The chip contains over 48,000 gene specific, 50 base pair probes representing the genome. Analysis of hybridization of RNA from experimental samples to these probes allows for quantification of genome wide expression. Briefly, the methodology involves, preparation of experimental samples, RNA extraction, in vitro synthesis of cRNA, biotin labelling of cRNA, hybridization of cRNA to the Illumina HT 12 bead chip and analysis of data (Figure 12).

This HT12 beadchip has been used successfully in many different contexts and diseases to analyse genome wide expression. Recently it was used to determine the mechanisms by which BRAF mutations affect tumorigenicity in colorectal cancer cell lines. Analysis of data obtained using the bead chip showed that BRAF inhibition results in loss of stemness and increased expression of differentiation markers such as *Caudal Type Homeobox1* and *2* (Boerries, Herr, Brummer, & Busch, 2015).

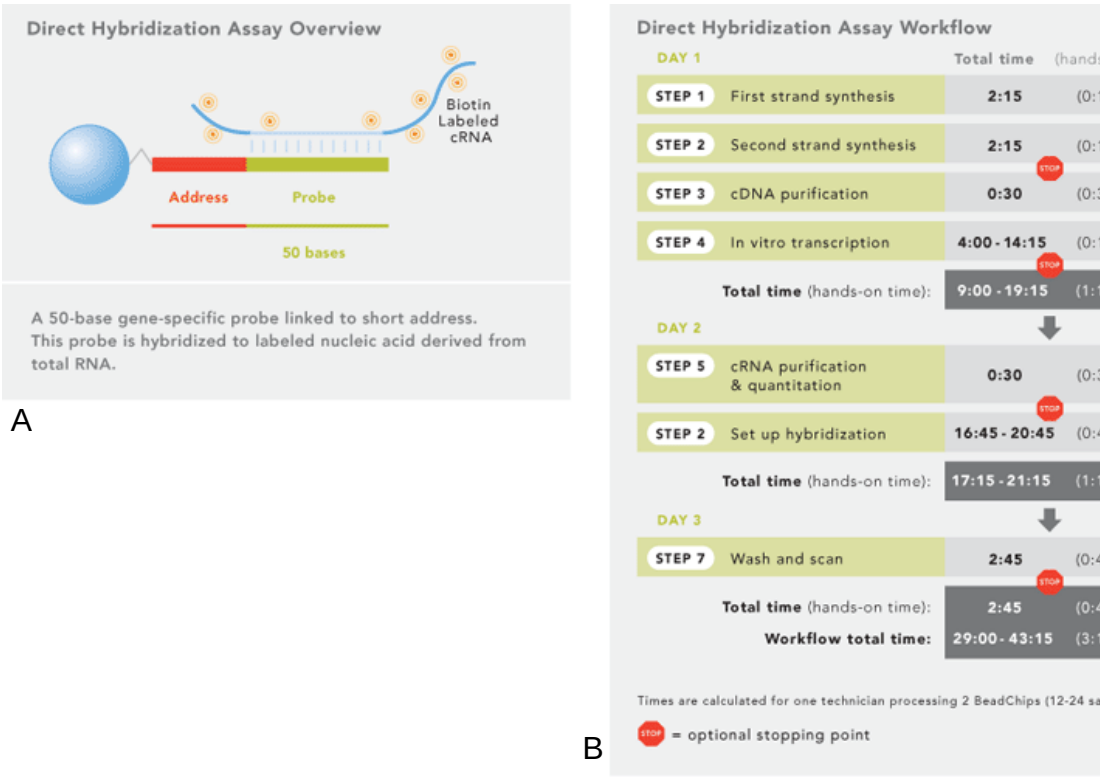


Figure 12 Illumina HT12 Beadchip
Figure adapted from http://www.illumina.com/technology/beadarray-technology/direct-hybridization-assay.html)
A) Schematic representation of a 50bp probe attached to a bead on the Illumina HT12 beadchip. Biotin labelled cRNA binds to probe and extent of hybridisation is quantified. B) Summary of the preparation and analysis steps for the Illumina HT12 beadchip.

3.2 Aims

There are two key aims to this chapter, the first was to analyse the effect of altering *TLE1* expression on genome wide expression. The second aim was to combine altering *TLE1* expression with stimulation of *NOD2* expression. These aims can be broken down as follows:

1. Optimise assays for altering (decreasing) *TLE1* expression in an appropriate cell line

2. Analyse the effect of knocking down *TLE1* on genome wide expression

3. Analyse the effect of stimulating *NOD2* expression on genome wide expression

4. Analyse the effect of combining *TLE1* knockdown with *NOD2* stimulation to identify any pathways in which both *TLE1* and *NOD2* function.

3.3 Results

3.4 Optimisation Experiments

3.4.1 Choice of cell line

Levels of *TLE1* were measured in four cell lines: HEK293, HCT116, HT29 and SW480 (Figure 13). Relative *TLE1* expression was standardised to expression in HCT116 cells. HEK293 cells were shown to have the highest level of *TLE1* expression, whereas SW480 cells had the lowest levels. Both HEK293 cells and SW480 cells have been used extensively in IBD research. Initially the SW480 cell line was chosen to be used in these experiments as MDP has been shown to stimulate *NOD2* expression in this cell line previously (Marian C Aldhous et al., 2011). My experiments showed transfection efficiencies for the control shRNA and *TLE1* shRNA were low and knockdown was undetectable by RT qPCR in SW480 cells. Hence, the HEK293 cell line was used in further experiments as it had the highest levels of *TLE1* (88% higher than SW480 cells) and transfection efficiency was much higher. Although it is not an intestinal epithelial cell line, it has been used in much IBD research (Nimmo et al., 2011; Stevens et al., 2013).

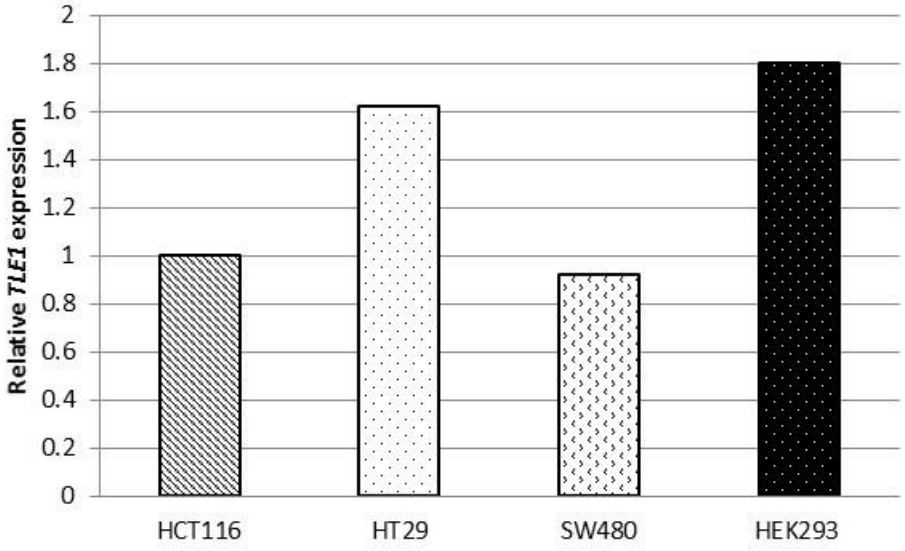


Figure 13 RT qPCR analysis of *TLE1* expression in HCT116, HT29, SW480 and HEK293 cells
TLE1 expression was analysed in three intestinal epithelial cell lines: HCT116, HT29, SW480 and one embryonic kidney cell line, HEK293. *TLE1* expression was quantified by RT qPCR and levels shown are relative to *TLE1* expression in HCT116 cells. This experiment was based on one replicate.

HEK293 cells were transfected with a mixture of four scrambled shRNA's (control) and four *TLE1* shRNA purchased from Genecopia. Cells were plated at 30% confluency in 6cm plates and transfected the following day. The total concentration of shRNA used per transfection was 2.5µg. Optimisation experiments were conducted at three different time points: 24, 48 and 72 hours, as recommended by the manufacturer. Levels of *TLE1* were measured by RT qPCR and normalised to *GAPDH* expression. Normalised *TLE1* levels at 24, 48 and 72 hours were compared in HEK293 cells transfected with control shRNA or *TLE1* shRNA (Figure 14). The 48 hour time point was most efficient for knockdown showing a 65% decrease in *TLE1* expression when compared to a cells transfected with a control construct for the same period of time. This was the time point used for cells in the final experiment.

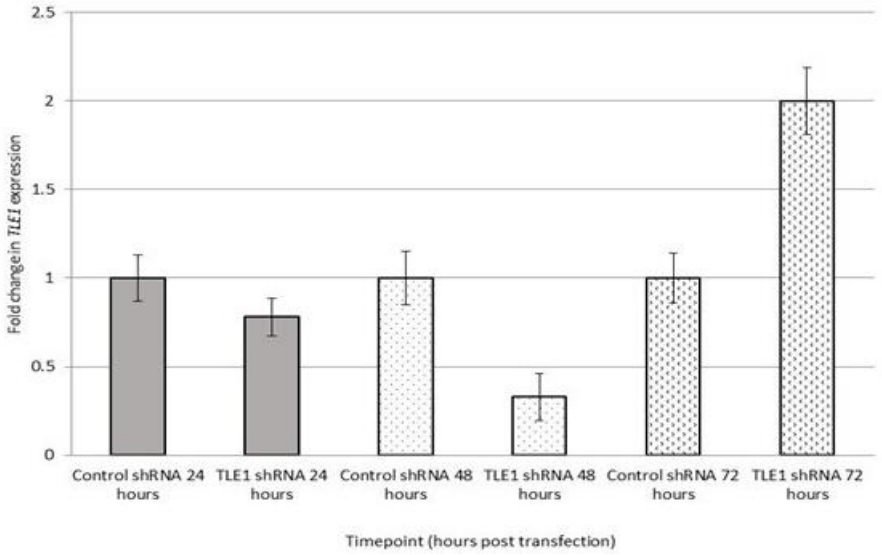


Figure 14 *TLE1* knockdown using shRNA is most effective 48 hours post transfection
TLE1 knockdown efficiency was analysed at 24, 48 and 72 hours post transfection of *TLE1* shRNA in HEK293 cells. Each time point is shown from left to right on the X axis. The Y axis shows fold change in *TLE1* expression, this is the fold change in *TLE1* expression between cells transfected with control shRNA construct and cells transfected with *TLE1* shRNA. Expression was quantified by RT qPCR. Each bar represents the average fold change in expression across three replicates. Error bars represent the standard error of the mean.

3.4.2 Optimisation of MDP mediated stimulation of *NOD2*

In these experiments, MDP was used to stimulate *NOD2* expression. Previous work has suggested that MDP stimulation of *NOD2* is most effective between 4 and 24 hours post stimulation (Marian C Aldhous et al., 2011). Figure 15 shows that MDP stimulation is highest at 24 hours post transfection, with a fold change of +2.51 (151% increase) between unstimulated and MDP stimulated cells.

A summary of the four different conditions (involving *TLE1* shRNA and MDP stimulation of *NOD2*) used in this experiment is shown in

Table 12.

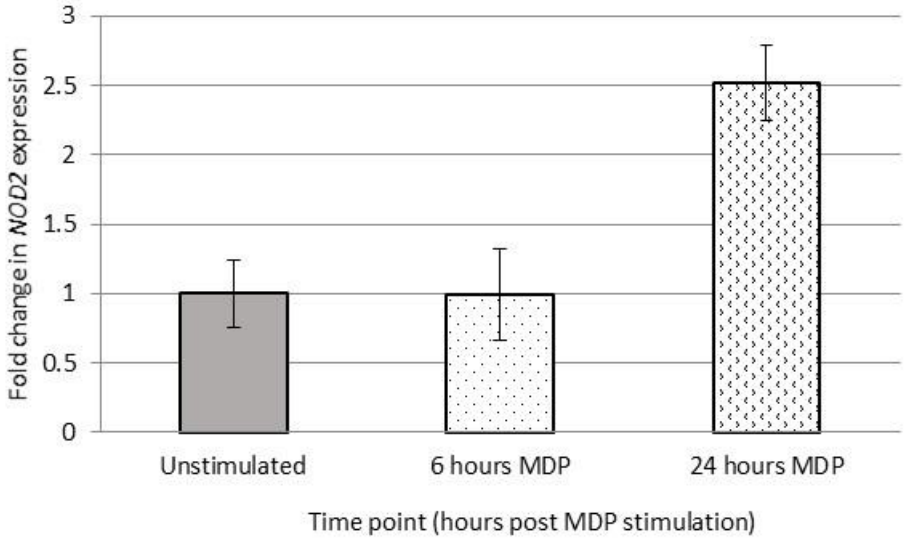


Figure 15 MDP stimulation of *NOD2* expression is most effective 24 hours post stimulation. MDP stimulation of *NOD2* expression was optimised in HEK293 cells at 6 and 24 hour time points. These time points were based on previous work in our laboratory (Marian C Aldhous et al., 2011). Fold change in expression was calculated by comparing unstimulated control cells and MDP stimulated cells. Expression was quantified by RT qPCR. Each bar represents the average fold change in *NOD2* expression across three replicates. Error bars represent the standard error of the mean.

	Unstimulated (24 hours)	MDP (*NOD2*) stimulated (24 hours)
Control shRNA (48 hours)	Control shRNA Unstimulated	Control shRNA MDP (*NOD2*) stimulated
TLE1 shRNA (48 hours)	*TLE1* shRNA Unstimulated	*TLE1* shRNA MDP (*NOD2*) stimulated

Table 12 Summary of experimental conditions used on Illumina HT12 Expression chip

The table shows the four experimental conditions analysed using the Ilumina HT12 Expression chip.

3.4.3 Reference gene optimisation for *TLE1* quantification

In order to determine the optimal reference genes to be used to quantify gene expression by RT qPCR, the stability of a panel of five reference genes was analysed. The reference genes used were *Ubiquitin C* (*UBC*), *Glyceraldehype-3-phosphate dehydrogenase* (*GAPDH*), *β Actin* (*β ACT*), *Tata box Binding Protein* (*TBP*) and *Succinate Dehydrogenase* (*SDHA*). These are well characterized reference genes that have been used previously in our laboratory and in other studies. Levels of each reference gene were analysed by RT qPCR. All optimisation conditions were analysed i.e. HEK293 cells transfected with: empty vector, empty vector and stimulated with MDP (6 and 24 hours), *TLE1* shRNA, *TLE1* shRNA and stimulated with MDP (6 and 24 hours). The qBase+ software (Biogazelle) which uses the geNorm algorithm was used to analyse the stability of each reference gene across all four samples (Vandesompele et al., 2002). The *GAPDH*, *UBC* and *βACT* reference genes were most stable and these were used in further RT qPCR experiments (Figure 16).

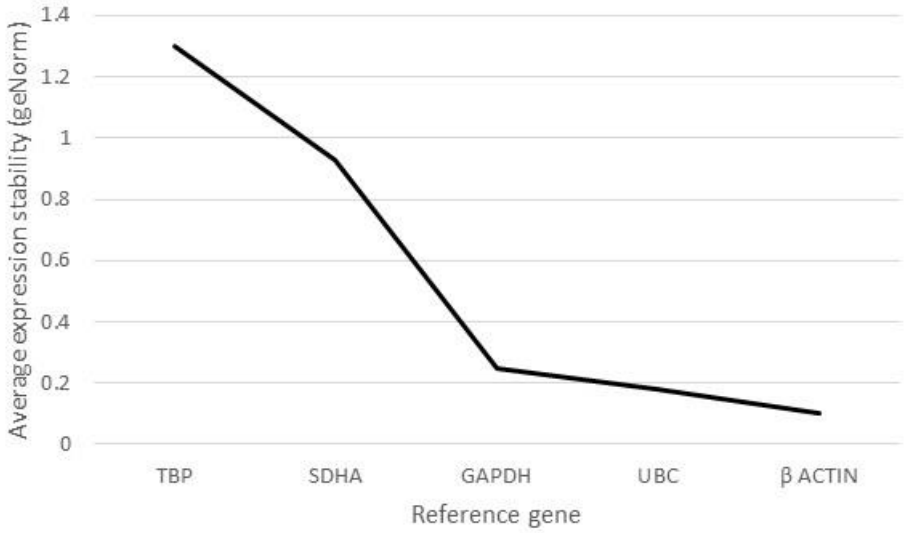


Figure 16 Average expression stability of RT qPCR reference genes

Line graph showing average expression stability of the five reference genes used in this project: *TBP* (*Tata box binding protein*), SDHA (*Succinate dehydrogenase), UBC* (Ubiquitin C) and *β Actin*. Average expression stability was calculated based on RT qPCR expression levels of all reference genes in HEK293 cells transfected with: empty vector construct, *TLE1* shRNA (24 and 48 hours), *TLE1* shRNA stimulated with MDP (6 and 24 hours) and empty vector construct stimulated with MDP (6 and 24 hours). GeNorm software was used to calculate average expression.

3.5 Final Experiment

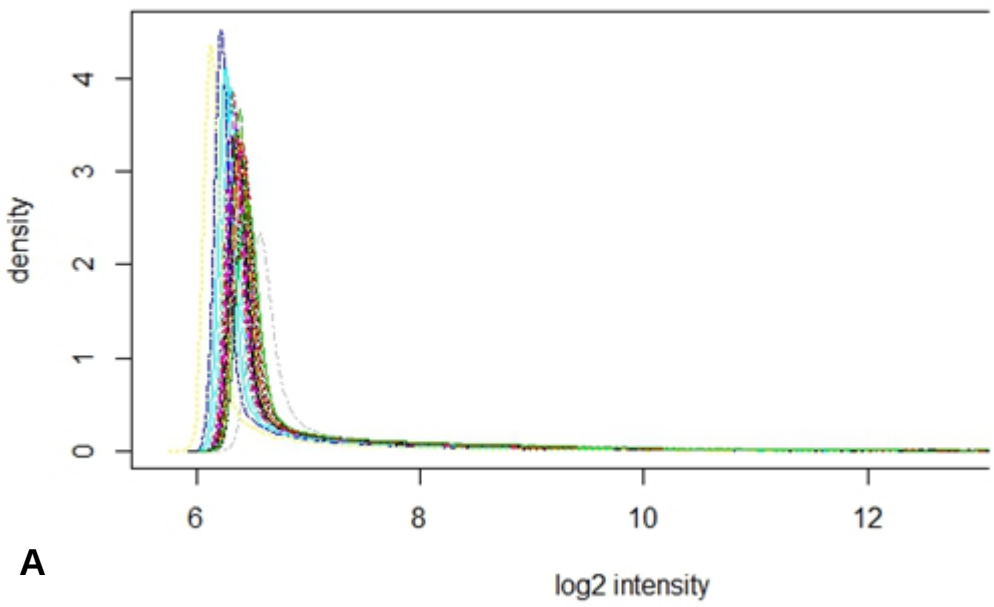
3.5.1 Data Analysis

In the final experiment, each of the optimised experimental conditions described in Table 12 was set up in triplicate and run on an Illumina HT12 expression chip. The experiment was performed by Dr Elaine Nimmo and the HT12 expression chip was run by the Welcome Trust Clinical Research Facility (Western General Hospital, Edinburgh). Data was analysed with the kind help of Dr Nicholas Kennedy and Dr Alex Adams using the Lumi and Limma packages in the statistical program, R.

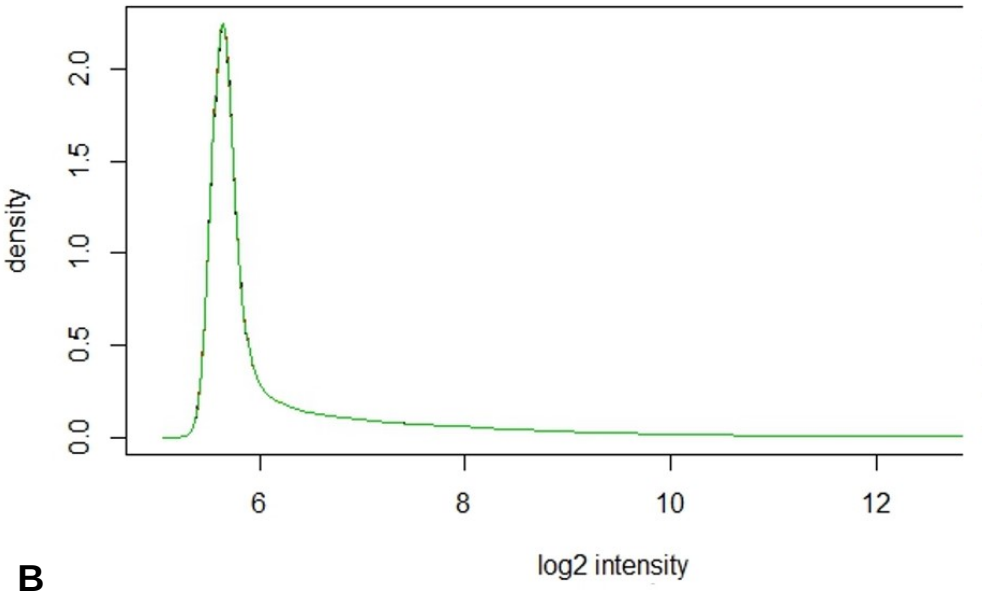
The first stage of data analysis involves normalising the data. Data is normalised in an attempt to compensate for the systematic differences between samples run across

different beadchips, to aid the identification of systematic biological differences between samples and eliminate any anomalies and spurious data trends. The data obtained from the HT12 expression chip was normalised using the VST transformation algorithm as part of the Lumi package in R. This method of normalisation was chosen based on evidence demonstrating that it improves detection of differentially expressed genes and reduces the number of false positive results when compared to other normalisation methods (Lin, Du, Huber, & Kibbe, 2008). It has been designed specifically for the Illumina platform and takes full advantage of the within array control probes (Lin et al., 2008).

Figure 17 shows the intensity of all the probes across the arrays, each line corresponds to a different sample number.Figure 17 (A) shows the density of the signal intensity across all the probes on the beadchip prenormalisation. Each line and therefore sample, is distinguishable, peak heights vary and the range of intensity also varies across samples. If this unnormalised data was used to anaylse the effect of *TLE1* knockdown we would expect many more spurious results due to variation between chips and array positions not being filtered out. Figure 17 (B) shows the data post normalisation, only one curve is visible as the samples all have equal ranges and intensities.



A



B

Figure 17 Log2 intensity of density of probes on the HT12 expression chip pre and post normalisation

Line graph showing density of ~47000 probes on the HT12 expression chip. There were four experimental samples and each experiment was conducted in triplicate. Each line on the graph represents one experimental replicate. A) shows the probe density across all unnormalised samples. B) Probe density across all normalised samples.

3.5.2 Confirmation of *TLE1* and *NOD2* expression by RT qPCR in samples used on Ilumina HT12 expression chip.

RT qPCR was used to determine whether *TLE1* knockdown and MDP stimulation of *NOD2* expression had been successful, all three technical replicates were analysed. Figure 18 shows that RT qPCR analysis revealed a 48% decrease in *TLE1* expression in cells transfected with *TLE1* shRNA compared to those transfected with control shRNA. Cells transfected with *TLE1* shRNA and stimulated with MDP showed a 43% decrease in expression when compared with those transfected with control shRNA and stimulated with MDP.

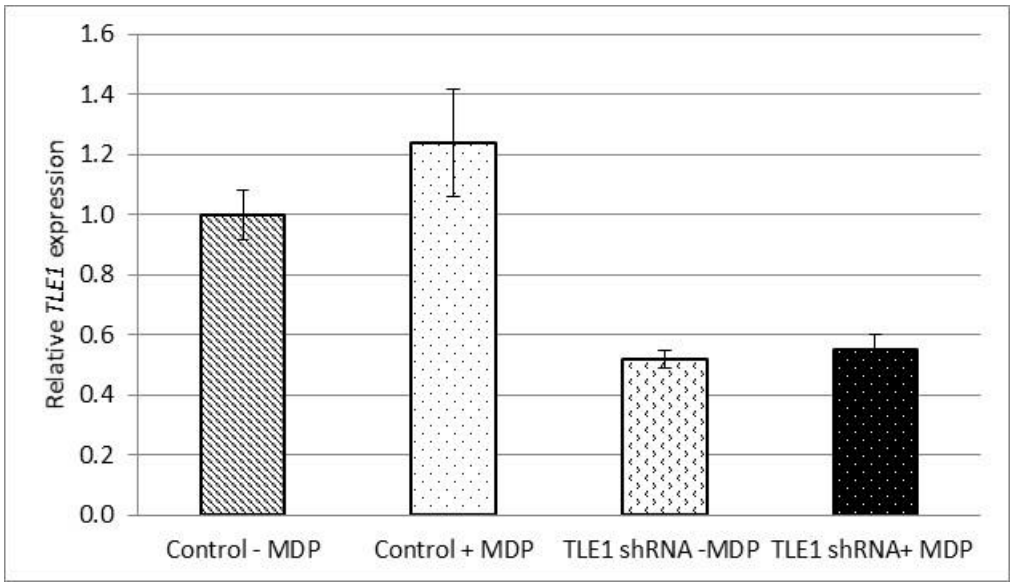


Figure 18 Confirmation of *TLE1* knockdown by RT qPCR in three technical replicates used on HT12 expression chip
RT qPCR analysis of *TLE1* expression across all three technical replicates of HEK293 cells transfected/ stimulated according to
Table 12 as per results from optimisation experiments (Figure 14 and Figure 15) Each bar represents average *TLE1* expression across three technical replicates. From left to right samples are: control shRNA (no MDP), control shRNA (with MDP), *TLE1* shRNA (no MDP), *TLE1* shRNA (with MDP). Error bars are calculated as ± SEM

Figure 19 shows RT qPCR analysis of *NOD2* expression in the same samples used in the expression chip. On average there was a 243% increase in *NOD2* expression in cells transfected with control shRNA and stimulated with MDP compared to

unstimulated cells. There was a lower average increase of 56% in *NOD2* expression in cells transfected with *TLE1* shRNA and stimulated with MDP when compared to untransfected controls. However, as shown by the error bars in Figure 19, there were large amounts of variation in *NOD2* expression levels between technical replicates of MDP stimulated cells. This variation makes this data very difficult to interpret and no definite conclusions can be drawn. This is further emphasised in Figure 21 which shows no detectable change in *NOD2* expression between unstimulated and stimulated samples according to results from the beadchip.

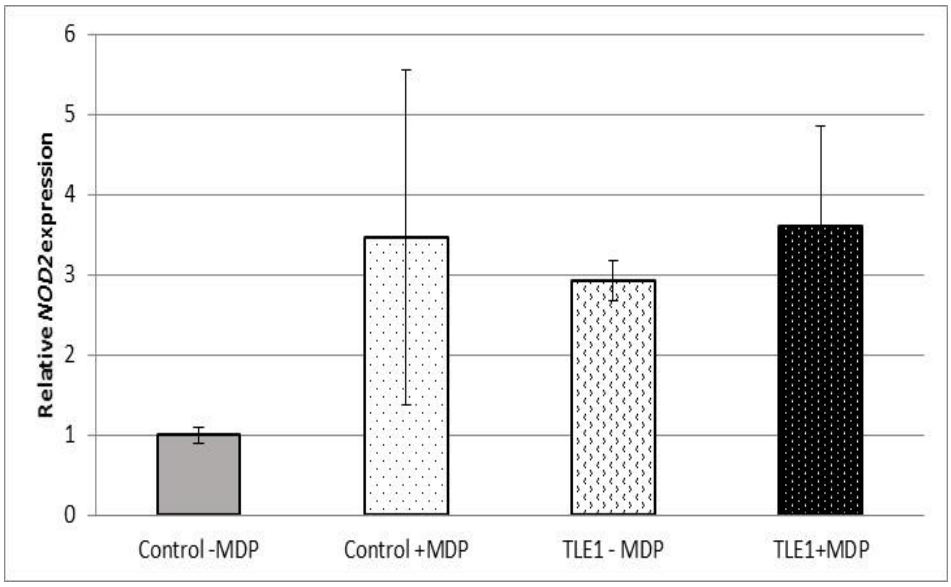


Figure 19 RT qPCR analysis of *NOD2* expression across the three technical replicates used on HT12 expression chip

Quantification of *NOD2* expression by RT qPCR on all three technical replicates of HEK293 cells transfected/ stimulated according to
Table 12, as per results from optimisation experiments (Figure 14 and Figure 15) Each bar represents the average *NOD2* expression across three technical replicates. From left to right samples are: control shRNA (no MDP), control shRNA (with MDP), *TLE1* shRNA (no MDP), *TLE1* shRNA (with MDP). Error bars are calculated as ± SEM.

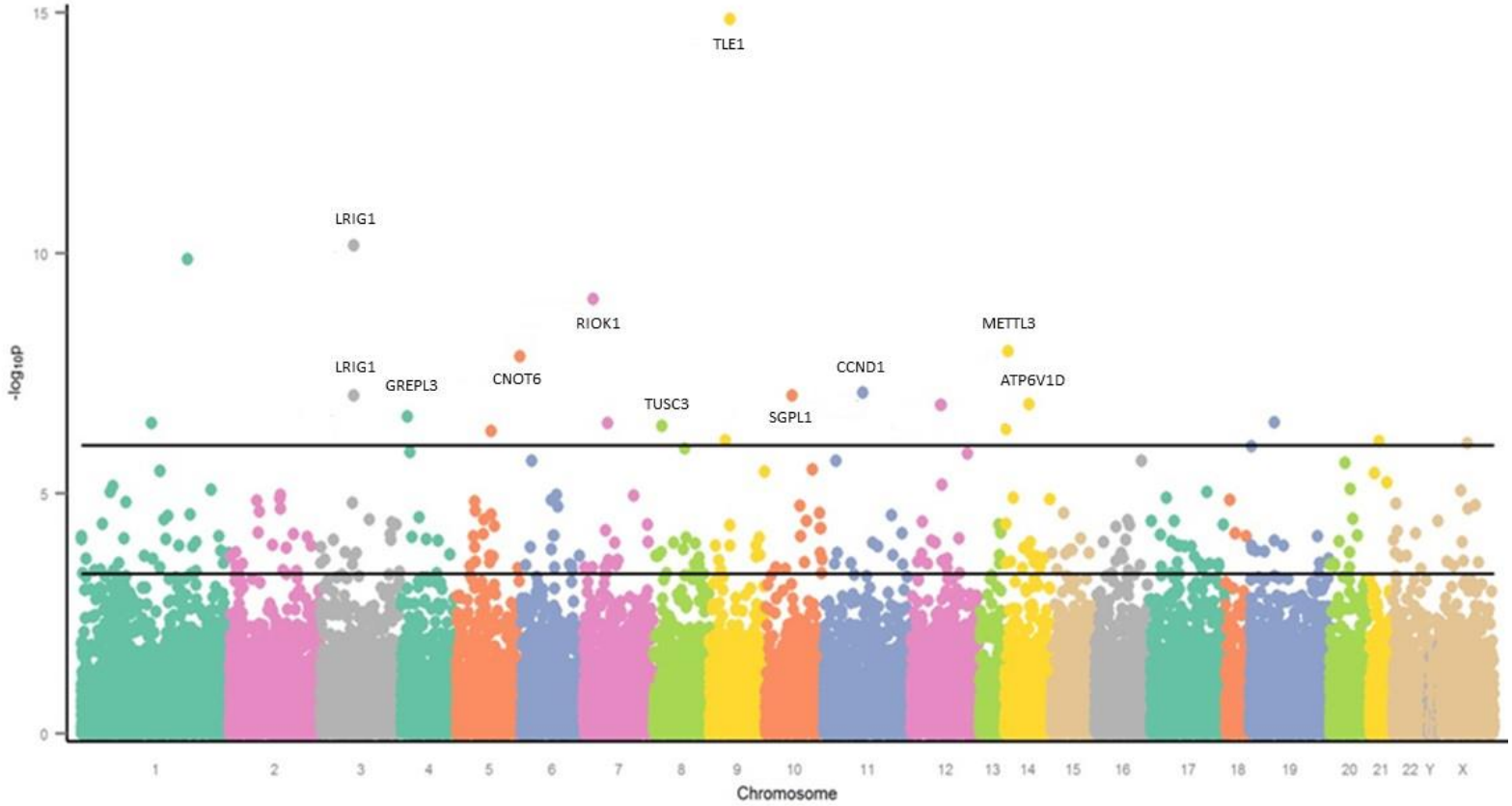


Figure 20 Differentially expressed genes following *TLE1* knockdown of HEK293 cells
Manhattan plot showing $-\log_{10}$ p values for all probes on the expression chip. Each colour along the X axis represents a different chromosome, chromosome number is shown below the X axis. P values shown relate to the differential expression between HEK293 cells transfected with an empty vector and *TLE1* shRNA. The first line indicates the threshold for False discovery rate (FDR) significance and the second denotes Bonferroni significance. Labelled genes reach the latter threshold.

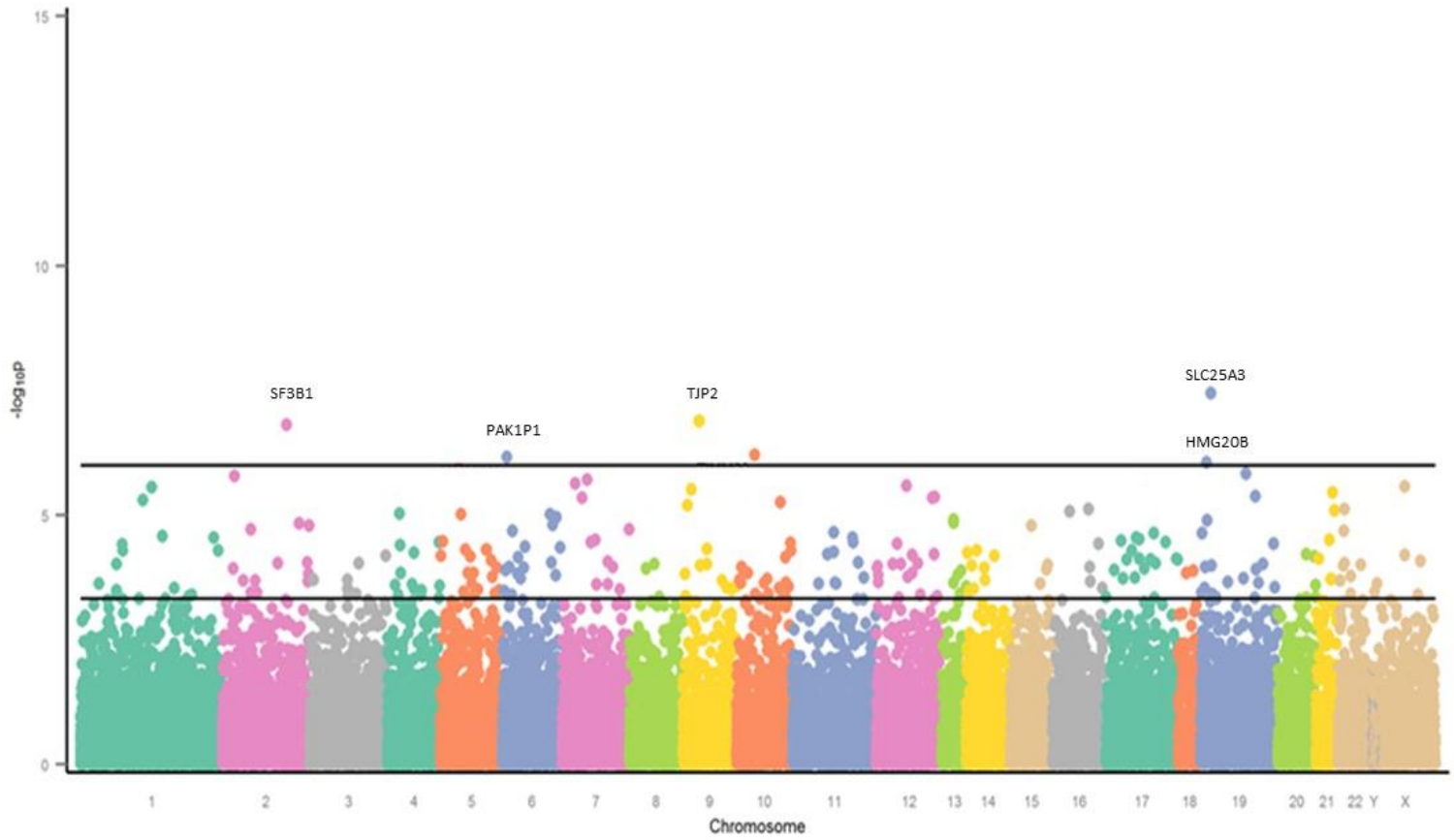


Figure 21 Differentially expressed genes following MDP stimulation of HEK293 cells
Manhattan plot showing $-\log_{10}$ p values for all probes on the expression chip following MDP stimulation. P values shown relate to differential expression between unstimulated and MDP stimulated HEK293 cells. Each colour along the X axis represents a different chromosome (chromosome number is shown below the X axis). The first line indicates the threshold for False discovery rate (FDR) significance and the second denotes Bonferroni significance. Labelled genes reach the latter threshold.

3.5.3 Differentially expressed genes following *TLE1* knockdown

TLE1 knockdown was successful as indicated by results from RT qPCR analysis (Figure 18) and results from the HT12 expression chip (Figure 20). Overall, 10 known genes reached Bonferroni significance and 526 genes FDR significance (Figure 20). Nine additional loci reached Bonferroni significance, however, actively transcribed regions have not been found near these loci. The focus of the following section was to analyse genes that were differentially expressed following *TLE1* knockdown according to the HT12 expression chip and identify any genes that may provide a link between *TLE1* and CD for follow up analysis. Differentially expressed genes were defined as genes with a Bonferroni corrected p value of less than 0.05. Identification of CD relevant genes was conducted by two separate means. A Pubmed search for each significant gene was carried out, with focus placed on identifying genes involved in inflammation, intestinal homeostasis, IBD, autoimmune disease, autophagy and bacterial exposure. Each gene was also input into the Search Tool for Interacting Genes (STRING) database. This database contains information about confirmed and predicted, direct and indirect protein interactions obtained from four sources, genomic context, high throughput and co expression experiments and data from Pubmed. All known IBD susceptibility loci were analysed for inclusion or proximity to Bonferroni corrected genes (Jostins et al., 2012).

Four genes were identified for follow up analysis: *SGPL1*, *TUSC3*, *RIOK1* and *CCND1*. Expression of each of these genes was analysed by RT qPCR following *TLE1* overexpression or *TLE1* knockdown in HEK293 cells. This was to confirm results from the beadchip and to further examine the relationship between each of these genes and *TLE1*. Each experiment was conducted in three biological replicates for both *TLE1* knockdown and overexpression. *TLE1* overexpression and downregulation was confirmed in each replicate, the Student's t test showed significant differences in expression between HEK293 cells overexpressing and down regulating *TLE1* (Figure 22) ($p=2.5 \times 10^{-3}$)

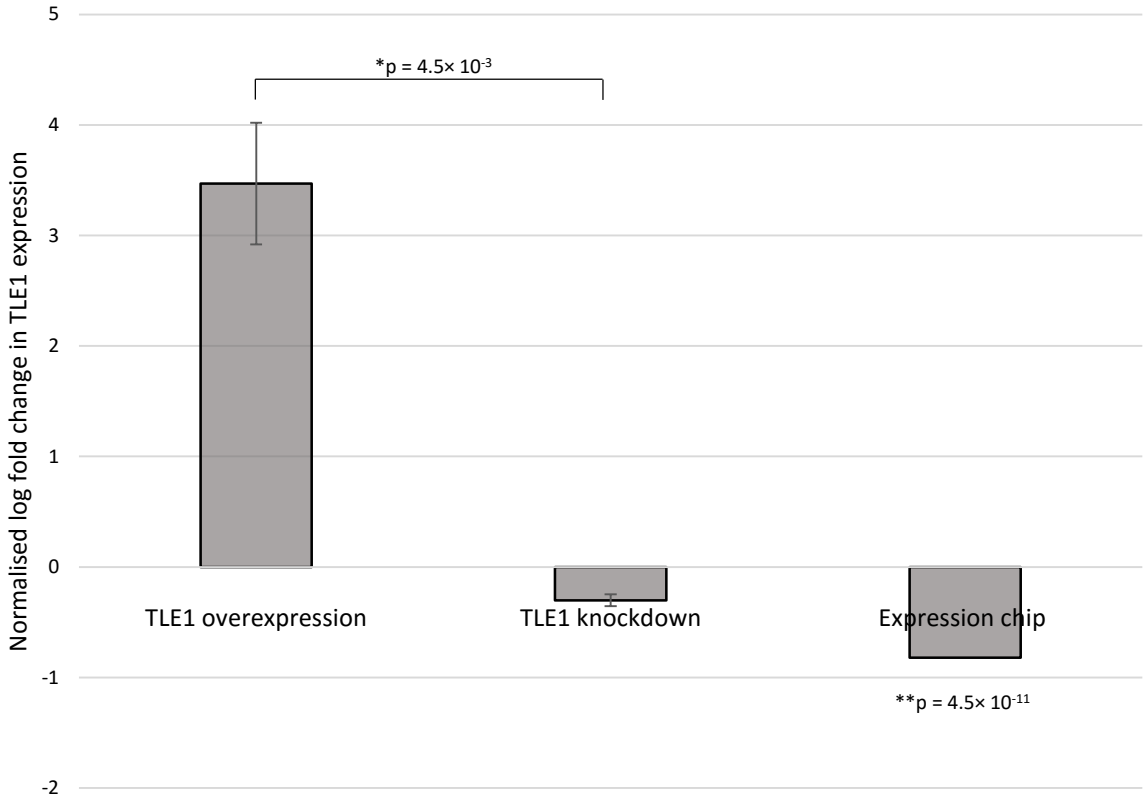


Figure 22 *TLE1* expression in cells with overexpressing and downregulating *TLE1*

RT qPCR analysis of *TLE1* overexpression and *TLE1* knockdown in three biological replicates. *TLE1* overexpression and *TLE1* knockdown bars represent average log fold change in expression. *TLE1* overexpression and knockdown data was normalised to expression of the reference genes: UBC, GAPDH and βACTIN. The bar labelled expression chip shows the log fold change in *RIOK1* expression as determined by the HT12 expression chip. Error bars represent the standard error of the mean. * p value as determined by Student's t test. ** p value as determined by Illumina HT12 expression chip.

3.5.4 RIO Kinase 1 (RIOK1)

RIO kinase 1 (*RIOK1*) is a serine/ threonine kinase that forms part of the protein arginine methyl transferase (PRMT) complex (Guderian et al., 2011). The PRMT complexes are involved in the post translational modification of proteins; they are function in the addition of methyl groups to arginine residues. This leads to epigenetic modifications of histone proteins and alteration of protein–protein interactions. RIOK1 is thought to be involved in the recruitment of substrate proteins for methylation by the PRMT complex. Proteins predicted to interact with it include a large number of

ribosomal proteins, implying a role in ribosomal biogenesis and RNA processing (Guderian et al., 2011). More specifically it has been implicated in 18S rRNA processing and cytoplasmic maturation of the 40S ribosomal subunit (Vanrobays, Gelugne, Gleizes, & Caizergues-ferrer, 2003; Widmann et al., 2012).

An IL-6 mediated role for ribosomal biogenesis has been implicated in UC, which suggests a possible mechanism by which RIOK1 may be involved in inflammation and IBD (Brighenti et al., 2014). Additional evidence suggesting a role for RIOK1 in IBD is provided by the STRING tool which predicts that RIOK1 interacts with the known IBD susceptibility proteins mTOR and VIMENTIN (Nimmo et al., 2011; Stevens et al., 2013). The interaction between RIOK1 and mTOR has been confirmed in HEK293T cells; the mTORC1 complex confirms interactions between RIOK1, RIOK2 and mTOR.

The PRMT complex has also been implicated in other autoimmune diseases and numerous cancers, including colorectal cancer (Y. Yang & Bedford, 2013; Zakrzewicz, Zakrzewicz, Preissner, Markart, & Wygrecka, 2012).

RIOK1 is a regulator of the PRMT complex and a by-product of protein arginine methylation is the production of ADMA (asymmetric dimethyl arginine). Increased ADMA levels have been attributed to defective PRMT complex and ADMA levels have been shown to be increased in IBD patients (Owczarek, Cibor, & Mach, 2010). This is a second mechanism by which decreased *TLE1* expression, which in turn downregulates *RIOK1* expression could cause accumulation/ deregulation of methylated proteins and increased ADMA. A similar pathogenic mechanism has been proposed in asthma (Zakrzewicz et al., 2012).

The Illumina HT12 expression chip showed significant down regulation of *RIOK1* upon *TLE1* knockdown. A log fold change of -0.12 was observed with a p value of 4.3×10^{-3}. RT qPCR analysis of *RIOK1* expression on three biological replicates of HEK293 cells down regulating *TLE1* confirmed this trend with an average log fold

change of -1.19 Overexpression of *TLE1* showed an average of +2.78 log fold change in *RIOK1* expression across three biological replicates (Figure 23).

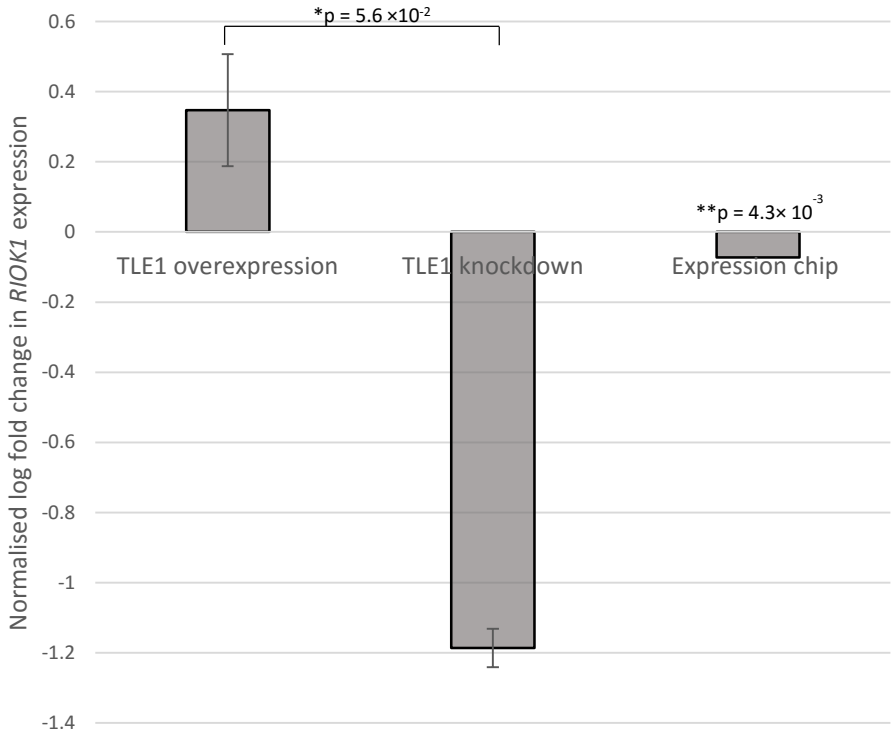


Figure 23 *RIOK1* expression in cells with overexpressing and downregulating *TLE1*

RT qPCR analysis of the effect of *TLE1* overexpression and *TLE1* knockdown on *RIOK1* expression across three biological replicates. *TLE1* overexpression and *TLE1* knockdown bars represent average log fold change in expression. *TLE1* overexpression and knockdown data was normalised to expression of the reference genes: UBC, GAPDH and β ACTIN. The bar labelled expression chip shows the log fold change in *RIOK1* expression as determined by the HT12 expression chip. Error bars represent the standard error of the mean. * p value as determined by Student's t test. ** p value as determined by Illumina HT12 expression chip.

RIOK1 expression in human intestinal biopsies

RIOK1 expression was analysed using RNA extracted from the terminal ileum of 53 CD patients and 31 healthy controls. Biopsies were taken from inflamed and non inflamed regions of each study participant. These experiments were conducted by Colin Noble as described in Noble et al, 2010. Briefly, total RNA from each biopsy

was extracted, converted to cRNA (complementary RNA) and purified, a reference cRNA genome was labelled with the fluorophore cyanine 3 (Cy3) and the test cRNA genome was labelled with cyanine 5 (Cy5), samples were then hybridized to an Agilent whole genome expression array, washed and then scanned for fluorescence. The microarray data was analysed with the help of Dr Nick Kennedy and Dr Alex Adams. *RIOK1* was a statistically significant marker of inflammation in healthy controls (repressed in inflamed tissues- p=0.007) but not in CD patients (p=0.907). There was a statistically significant increase in *RIOK1* expression between non inflamed controls and CD patients (p=0.02) and a statistically significant decrease in average *RIOK1* expression between inflamed healthy controls and CD patients (p=0.003) (Figure 24). In summary, *RIOK1* is a marker for inflammation in healthy controls but not CD patients and is differentially expressed between both inflamed and non inflamed CD patients and controls, however the change seen in between inflamed patients and controls is in the opposite direction to that seen in no inflamed patients and controls.

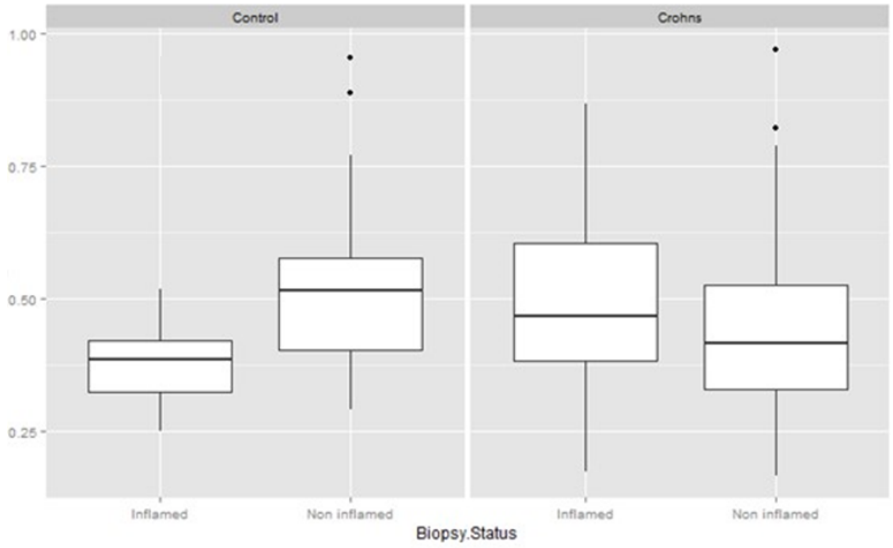


Figure 24 *RIOK1* mRNA expression in human intestinal biopsies

RIOK1 expression in human inflamed and non inflamed intestinal biopsies from healthy controls and CD patients. Sample collection, experiment and data analysis were performed by Colin Noble (Noble et al., 2010). Boxplots describe average *RIOK1* expression across the 53 CD patients and 31 controls used by Noble *et al*. Genome wide expression was analysed using Agilent whole genome microarrays. Bonferroni corrected p values represent differences between inflamed and non inflamed individuals in controls and CD patients. Black circles represent outliers.

3.5.5 Cyclin D1 (CCND1)

CCND1 encodes CYCLIN D1, a protein involved in cell cycle regulation. The cyclins are a group of proteins involved in progression of a cell from the G1 to the S phase of the cell cycle. Cyclins function via interaction with their catalytic counterparts the cyclin dependent kinases (cdk's), CYCLIN D1 interacts with CDKs 4 and 6. Binding of cyclins to cdk's results in phosphorylation of downstream signalling molecules including the retinoblastoma (Rb) protein. Rb is a cell cycle inhibitor, which is inhibited upon phosphorylation. This leads to initiation of DNA replication and movement from G1 to the S phase of the cell cycle. Hence cyclins play a key role in the promotion of cell proliferation.

Recent evidence suggests that CYCLIN D1 also has cdk independent functions. It interacts with over 30 different transcription factors and co factors. These transcriptional regulators are involved in cell differentiation, apoptosis, cell migration, metabolism (mitochondrial function) and cell growth.

CCND1 expression is induced by growth factors including: TGFβ, various hormones e.g. gastrin in the GI tract, IGF1 and II and a variety of other extracellular signalling molecules. The promoter region of *CCND1* contains binding sites for a number of known oncogenic proteins such as RAS, β CATENIN and STATs.

CCND1 was identified as a gene of particular interest to study due to, its functional relevance to IBD and interactions with proteins implicated in IBD pathogenesis and evidence for its role in IBD. STRING analysis showed that CCND1 is predicted to interact with TLE1 and MTOR. Mechanistic target of rapamycin (MToR) is a serine/threonine kinase that is activated in response to cellular stress. It regulates translation of specific mRNAs via interaction with transcription factors. mTOR inhibitors such as rapamycin are being explored as IBD therapeutics, as they have immunosuppressant effects in patients with IBD. (Massey, Bredin, & Parkes, 2008)

The intestinal epithelium is one of the most rapidly proliferating tissues of the body and maintenance of mucosal homeostasis is dependent on precise control of cell proliferation in the intestinal crypts. As previously discussed CYCLIN D1 is essential for cell division and the progression of a cell from the G1 to S phase of the cell cycle, hence it is crucial to intestinal homeostasis. *CCND1* is highly expressed at sites of inflammation; in IBD models, *CCND1* has been shown to be upregulated in both epithelial and immune cells (Taylor, 2006). In line with this, immunhistochemical analysis of intestinal biopsies has revealed increased expression of CYCLIN D1 in IBD (both UC and CD) (Ioachim, Michael, & Agnantis, 2004). There is also some evidence for a contribution of *CCND1* to IBD related colorectal carcinoma, however the evidence is mixed (Dekken et al., 2007; Ioachim et al., 2004; Kanaan et al., 2011).

Given the dysregulation of *CCND1* expression in IBD, it has been the target of therapeutic strategies for IBD. The use of siRNA to silence cyclin D1 expression in the gut specific leukocytes of mouse models has shown down regulation of *CCND1* can reverse DSS induced colitis (Peer, Park, Morishita, Carman, & Shimaoka, 2008). More recent work has confirmed the therapeutic potential of cyclin D1 inhibitors in IBD treatment (Kriegel & Amiji, 2011).

The Illumina HT12 expression chip showed significant up regulation of *CCDN1* upon *TLE1* knockdown. A +0.54 log fold change was observed with a p value of 2.7×10^{-3}. Analysis of *CCND1* expression on three biological replicates of HEK293 cells down regulating *TLE1* confirmed this trend with an average log fold change of +0.63. Overexpression of *TLE1* showed an average of +0.42 log fold change in *CCDN1* expression across three biological replicates (Figure 25).

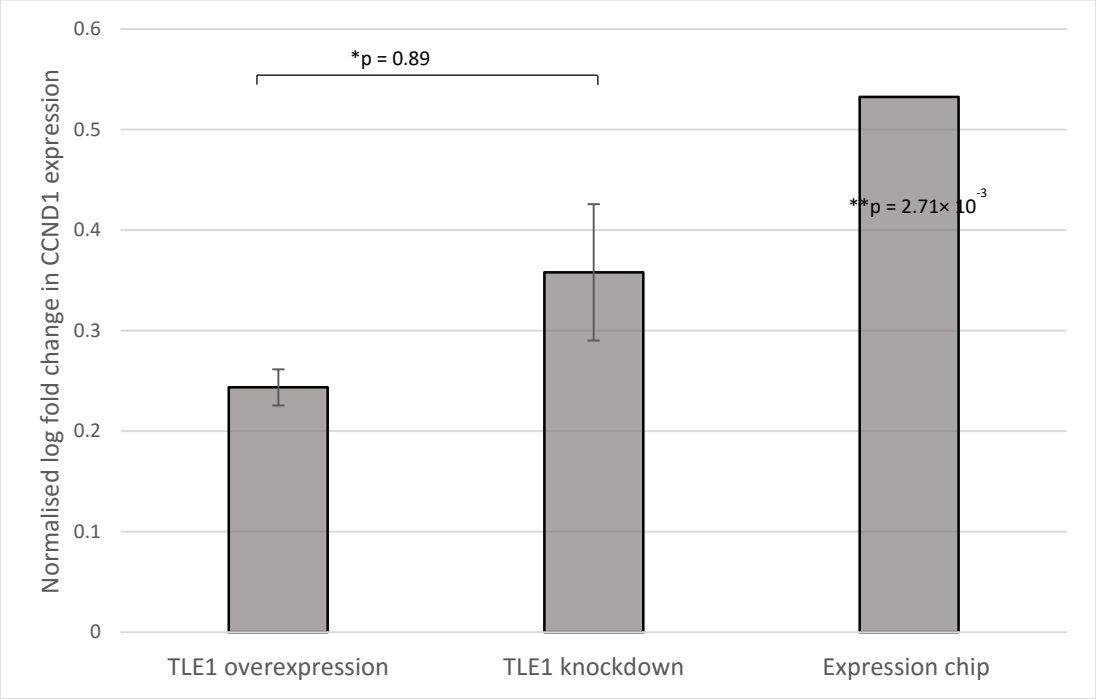


Figure 25 *CCND1* expression in cells with overexpressing and downregulating *TLE1*

RT qPCR analysis of the effect of *TLE1* overexpression and *TLE1* knockdown on *CCND1* expression across three biological replicates. *TLE1* overexpression and *TLE1* knockdown bars represent average log fold change in expression. *TLE1* overexpression and knockdown data was normalised to expression of the reference genes: UBC, GAPDH and βACTIN. The bar labelled expression chip shows the log fold change in *CCND1* expression as determined by the HT12 expression chip. Error bars represent the standard error of the mean. * P value as determined by Student's t test. ** P value as determined by Illumina HT12 expression chip.

CCND1 mRNA expression in human intestinal biopsies

CCND1 was a statistically significant marker of inflammation in CD patients (p=0.007), the same trend is seen in healthy controls although the p value does not reach significance (p=0.17), this is likely to be due to small sample sizes. There was no difference in *CCND1* expression between inflamed healthy controls and CD patients (p=0.16) or non inflamed healthy controls and CD patients (p=0.80) (Figure 26). In summary, *CCND1* seems to be a marker for inflammation in CD patients and healthy controls and is not differentially expressed between healthy controls and CD patients regardless of inflammation status (Figure 26).

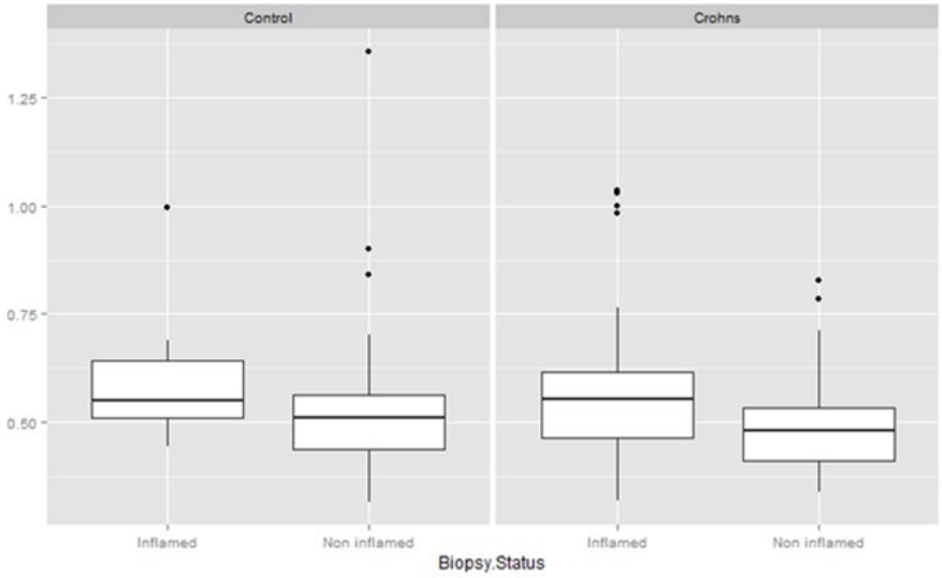


Figure 26 *CCND1* mRNA expression levels in human intestinal biopsies

CCND1 expression in human inflamed and non inflamed intestinal biopsies from healthy controls and CD patients. Sample collection, experiment and data analysis were performed by Colin Noble (Noble et al., 2010). Boxplots describe average *CCND1* expression across the 53 CD patients and 31 controls used by Noble *et al*. Genome wide expression was analysed using Agilent whole genome microarrays Bonferroni corrected p values represent differences between inflamed and non inflamed individuals in controls and CD patients. Black circles represent outliers.

3.5.6 Sphingosine 1 Phosphate Lyase (SGPL1)

SGPL1 is a ubiquitously expressed enzyme which is responsible for the conversion of sphingosine 1 phosphate (S1P) to 2-hexadecenal and phosphoethanolamine. Hence it is thought to be essential for homeostasis of S1P levels. S1P accumulation causes normal cells to be transformed into cancerous cells. S1P is a lipid signalling molecule which binds to the G protein coupled receptors: $S1P_1$, $S1P_2$, $S1P_3$, $S1P_4$ and $S1P_5$. Binding of S1P to $S1P_1$-$S1P_6$ results in inhibition of apoptosis, increased angiogenesis and amplified NFκB and STAT3 signaling.

The JAK/STAT signalling pathway has previously been implicated in IBD. Polymorphisms in members of the pathway including signal transducing and activator

of transcription (STAT3) (Barrett et al., 2008) The STAT proteins are cytoplasmic transcription factors that are activated following phosphorylation of tyrosine residues. STAT3 activation leads to release of pro inflammatory cytokines including Il-6. Activation of STAT3 and release of pro inflammatory cytokines results in differentiation of Th17 cells. S1P signalling is part of a reciprocal positive feedback loop with STAT3, this signalling is thought to play a crucial role in inflammation and cancer, and could also provide a mechanism by which TLE1/SGPL1 influence CD susceptibility.

S1P signaling is particularly relevant to IBD as epithelial cells in the small intestine are exposed to the breakdown products of sphingolipids; for example S1P is broken down by SGPL1 to 2-hexadecenal and phosphoethanolamine, these products would be exposed to the intestinal epithelium. In normal, healthy gut tissues sphingosine molecules enter the gut and are phosphorylated to S1P and then S1P is rapidly degraded by SGPL1. In this normal tissue rapid degradation is dependent upon high expression of SGPL1 in enterocytes.

The Illumina HT12 expression chip showed significant up regulation of *SGPL1* upon *TLE1* knockdown. A log fold change of +0.48 was observed with a p value of 4.3×10^{-3}. RT qPCR analysis of *CCND1* expression on three biological replicates of HEK293 cells down regulating *TLE1* confirmed this trend with an average log fold change of +0.47. Overexpression of *TLE1* showed an average of +1.83 log fold change in *SGPL1* expression across three biological replicates (Figure 27).

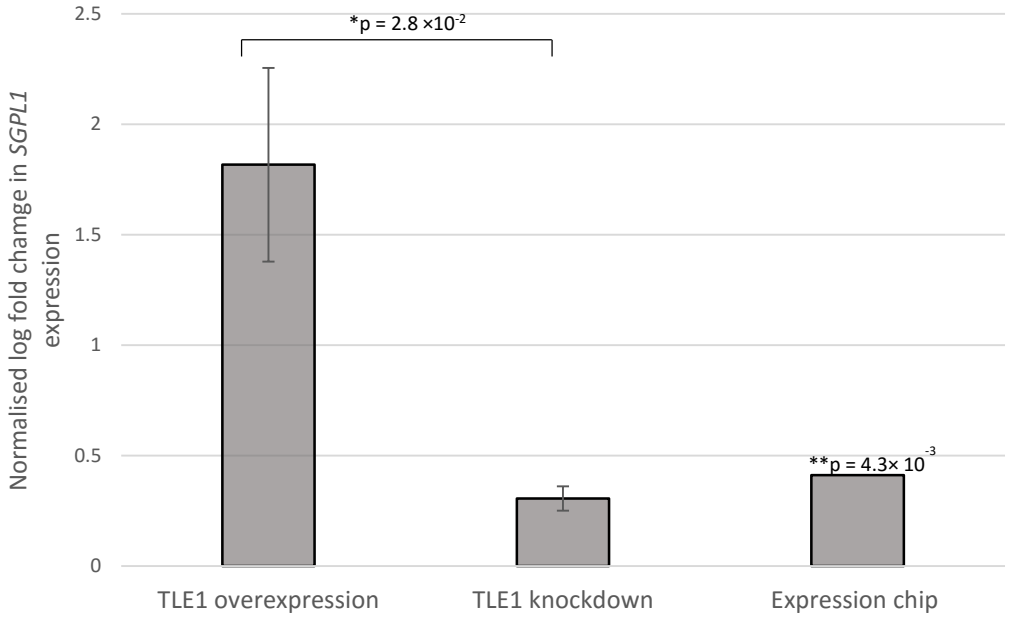


Figure 27 *SGPL1* expression in cells with overexpressing and downregulating *TLE1*

RT qPCR analysis of the effect of *TLE1* overexpression and *TLE1* knockdown on *SGPL1* expression across three biological replicates. *TLE1* overexpression and *TLE1* knockdown bars represent average log fold change in expression. *TLE1* overexpression and knockdown data was normalised to expression of the reference genes: UBC, GAPDH and βACTIN. The bar labelled expression chip shows the log fold change in *SGPL1* expression as determined by the HT12 expression chip. Error bars represent the standard error of the mean. * P value as determined by Student's t test. ** P value as determined by Illumina HT12 expression chip.

SGPL1 mRNA expression in human intestinal biopsies

SGPL1 is a marker for inflammation in CD patients (repressed in inflamed tissue p=0.04) but not in healthy controls (p=0.98). There was no difference in *SGPL1* expression between inflamed healthy controls and CD patients (p=0.98) or non inflamed healthy controls and CD patients (p=0.14) (Figure 28). In summary, *SGPL1* is a marker for inflammation in CD patients but not healthy controls and is not differentially expressed between healthy controls and CD patients regardless of inflammation status (Figure 28).

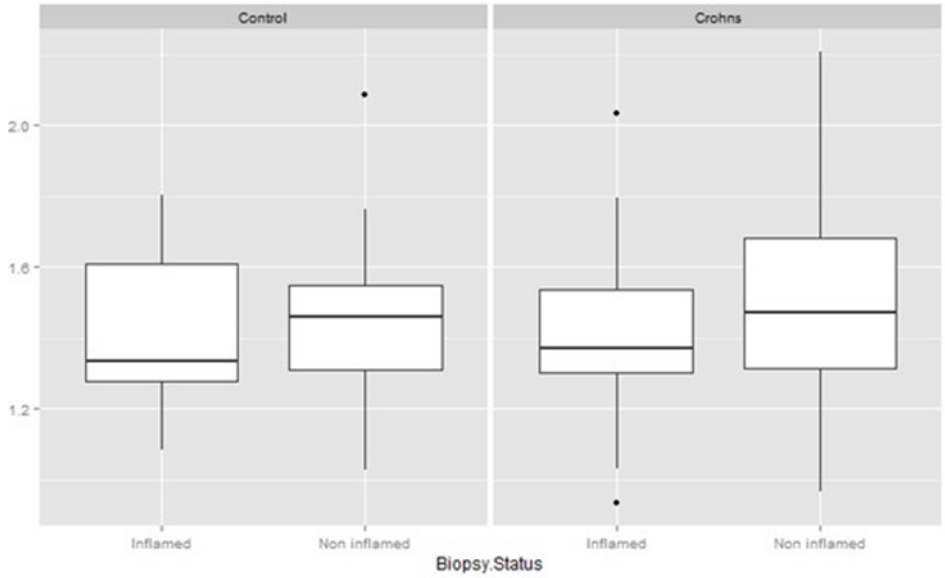


Figure 28 SGPL1 mRNA expression in human intestinal biopsies

SGPL1 expression in human inflamed and non inflamed intestinal biopsies from healthy controls and CD patients. Sample collection, experiment and data analysis were performed by Dr Colin Noble, Dr Nick Kennedy and Dr Alex Adams (Noble et al., 2010). Boxplots describe average *SGPL1* expression across the 53 CD patients and 31 controls used by Noble *et al*. Genome wide expression was analysed using Agilent whole genome microarrays. Bonferroni corrected p values represent differences between inflamed and non inflamed individuals in controls and CD patients. Black circles represent outliers.

3.5.7 Tumour suppressor candidate 3 (TUSC3)

Tumour suppressor candidate 3 (*TUSC3*) is the human homologue of the yeast *ostp3* gene which functions as part of the oligosaccharyltransferase complex (OST). The human OST complex is composed of seven different subunits, one of which is TUSC3. It has been shown that *TUSC3* is present in the rough ER (RER) and binds to core proteins of the OST complex: STT3A and STT3B in ovarian cancer and kidney cell lines respectively (Horak et al., 2014; Vaňhara et al., 2013).

The OST complex is essential for N-linked glycosylation of proteins in the ER. N linked glycosylation is an essential post translational modification of eukaryotic proteins that moderates protein folding and structure, stops proteins from being

degraded and alters function and immunogenicity. Defects in N linked glycosylation can lead to the accumulation of unfolded or misfolded proteins in the ER. This causes ER stress and initiates the unfolded protein response (UPR). The UPR is composed of three pathways: the Ire1α pathway, the ATF6α pathway and the PERK signalling pathway (Figure 29). As discussed in (see Chapter 4), proteins in all three UPR pathways have been implicated in IBD and ER dysfunction has been shown to influence IEC function and IBD pathogenesis. Dysregulation of *TUSC3* alters the ER structure and the ER stress response (Horak et al., 2014; Kratochvílová et al., 2015). The ER chaperone, Binding immunoglobulin protein (BiP) and the transcription factor C/EBP homologous protein (CHOP), which are essential to the PERK, IRE1α, ATF6 and PERK UPR pathways respectively, are down regulated in response to *TUSC3* silencing (Kratochvílová et al., 2015). Down regulation of *TUSC3* by shRNA leads to reduced expression of Ire1α.

The most widely accepted environmental IBD risk factor is smoking; it has been shown to affect both IBD incidence and history. *TUSC3* methylation has been shown to decrease in smokers (Imboden et al., 2012). Alteration of methylation at this locus could disrupt expression of the gene or alter binding of transcriptional regulators thereby influencing susceptibility to CD.

The Illumina HT12 expression chip showed significant up regulation of *TUSC3* upon *TLE1* knockdown. A log fold change of -0.44 was observed with a p value of 1.8×10^{-2}. RT qPCR analysis of *TUSC3* expression on three biological replicates of HEK293 cells down regulating *TLE1* confirmed this trend with an average log fold change of -1.42. Overexpression of *TLE1* showed an average of -0.73 log fold change in *TUSC3* expression across three biological replicates (Figure 30).

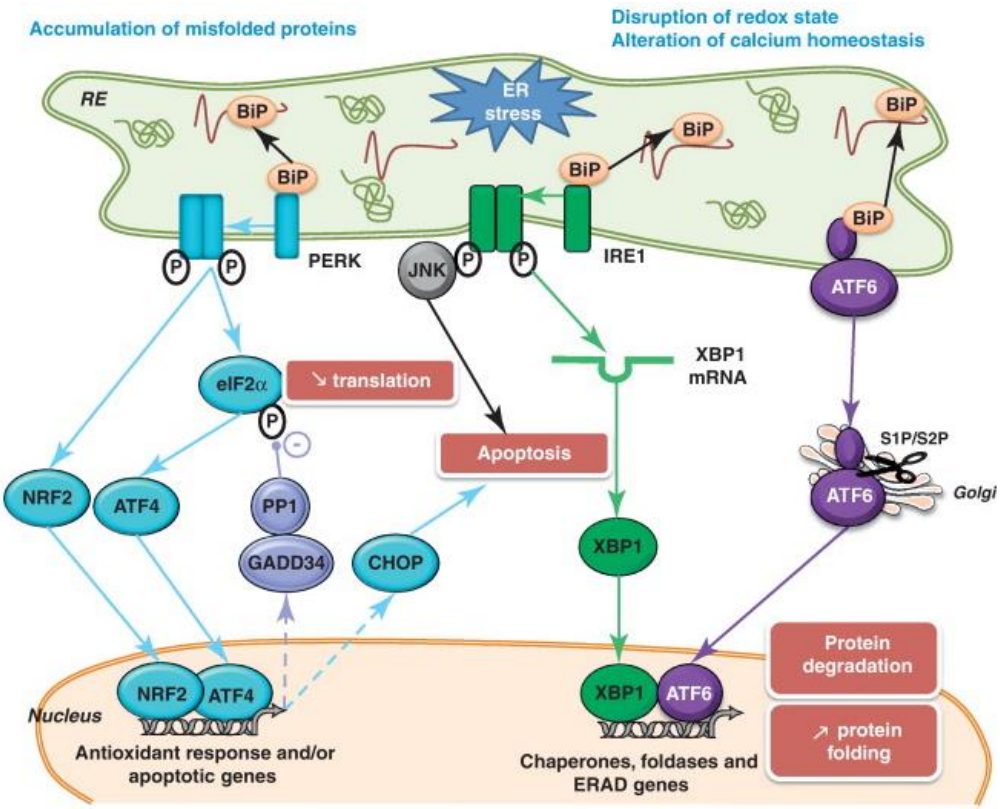


Figure 29 Summary of UPR signalling pathways (Wu & Kaufman, 2006)

Schematic diagram showing the sensor, transducer and effector proteins of the three UPR pathways: PERK, Ire1α and ATF6.

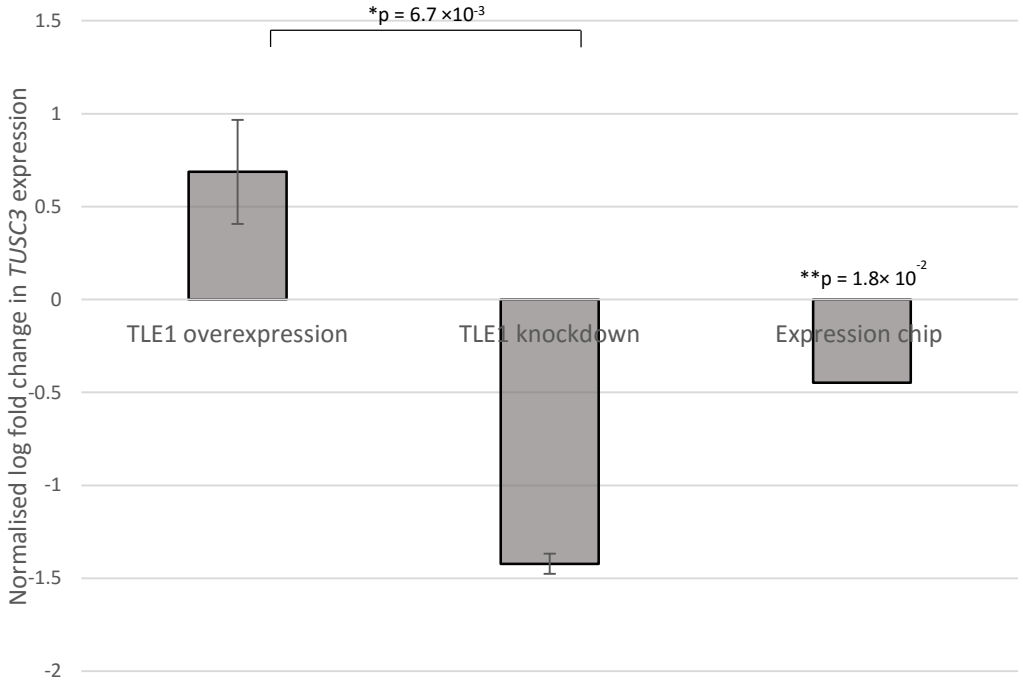


Figure 30 *TUSC3* expression in cells with overexpressing and downregulating *TLE1*

RT qPCR analysis of the effect of *TLE1* overexpression and *TLE1* knockdown on *TUSC3* expression across three biological replicates. *TLE1* overexpression and *TLE1* knockdown bars represent average log fold change in expression. *TLE1* overexpression and knockdown data was normalised to expression of the reference genes: UBC, GAPDH and β ACTIN. The bar labelled expression chip shows the log fold change in *TUSC3* expression as determined by the HT12 expression chip. Error bars represent the standard error of the mean. * P value as determined by Student's t test. ** P value as determined by Illumina HT12 expression chip.

TUSC3 mRNA expression in human intestinal biopsies

Data for *TUSC3* expression was not available, two *TUSC3* probes had been removed from the analysis as they did not meet quality control criteria.

3.5.8 Differentially expressed genes following MDP stimulation of *NOD2* expression

Although the average levels of *NOD2* expression were increased upon MDP stimulation, confidence limits imply that the difference was not statistically significant. Additionally *NOD2* did not reach statistical significance and was not detectably differentially expressed in MDP stimulated versus unstimulated cells as assessed by the HT12 expression chip. This may be because *NOD2* levels in these cell lines are low and below the levels of detection on the HT12 expression chip. In line with this, no further pathway analysis or follow up of other differentially expressed genes was carried out on this data set.

3.6 Discussion

In this chapter the effects of *TLE1* knock down in HEK293 cells, on genome wide expression were analysed using an Illumina HT12 expression chip. Of the 10 known loci that were differentially expressed, 5 were identified as particularly relevant to IBD pathogenesis: *RIOK1*, *CCND1*, *SGPL1* and *TUSC3*. Expression of each of these genes was then analysed at the mRNA level in ileal biopsies from 53 CD patients and 31 controls.

It is important to note that the expression data from ileal biopsies has a number of limitations and hence the p values presented should be interpreted with caution. The healthy control and CD patient group sizes were both relatively small, 53 and 31 respectively and were not identical in size. Furthermore, although age matching was attempted for the samples when conducting the study, all samples are not age matched which may hinder interpretation of the results. Finally, of particular relevance to the genes identified as markers of inflammation in healthy controls, the inflamed healthy controls may have had another inflammatory disease, resulting in altered expression of the genes in question. The biopsy expression analysis results shown in this chapter can be interpreted as preliminary results and may be the focus of future research in larger Scottish cohorts.

RIOK1 is a serine/threonine kinase that is involved in the post translational methylation of arginine residues. It has been implicated in ribosomal biogenesis and is predicted to interact with VIMENTIN and mTOR, both of which have been implicated in IBD. Overexpression of *TLE1* lead to an increase in *RIOK1* expression whereas *TLE1* knockdown had the opposite effect, causing a decrease in *RIOK1* levels. This is interesting as *TLE1* is transcriptional co repressor and you would expect the opposite effect. However, there is some evidence that *TLE1* may also behave as a transcriptional activator. TLE1 has been shown to interact with Estrogen Related Receptor γ and when co expressed in kidney cells TLE1 results in increased expression of *ERRγ* as determined by a luciferase assay (Hentschke & Borgmeyer, 2003). In addition to

RIOK1 being differentially expressed upon *TLE1* knockdown ($p=2.71\times10^{3}$) the difference between *RIOK1* mRNA expression between cells over expressing *TLE1* and cells down regulating *TLE1* was statistically significant ($p=5.6\times10^{-2}$). *RIOK1* appears to be a marker for inflammation for CD patients but not healthy controls, implying a CD specific role in inflammation. However it is not differentially expressed between healthy controls and CD patients regardless of inflammation status.

CCND1 is a cell cycle regulator that is involved in inflammation and has been shown to be over expressed in intestinal biopsies from IBD patients, *CCND1* siRNA has been shown to have therapeutic potential in mice (Ioachim et al., 2004; Kriegel & Amiji, 2011). Both overexpression and knockdown of *TLE1* led to an increase in *CCND1* expression, this difference was not statistically significant (p=0.89). Additionally *CCND1* mRNA was not differentially expressed between healthy controls and CD patients (p= 0.89). The latter is in contrast to results from other studies which analysed CCND1 levels by immunohistochemistry of human intestinal biopsies this could be attributable to the differences in methodology between the two studies (Ioachim et al., 2004).

SGPL1 is an enzyme which functions as part of S1P signalling and in turn JAK/STAT signalling. S1P signalling interplays with JAK/STAT signalling, members of the JAK/STAT pathway including STAT3 have been implicated in IBD (Barrett et al., 2008). *SGPL1* mRNA expression was increased upon both *TLE1* knock down and overexpression, this may be due to additional *SGPL1* regulators that are compensating for *TLE1* dysregulation. *SGPL1* was differentially expressed between inflamed controls and CD patients but not their non- inflamed counterparts. It is also differentially expressed between inflamed and non-inflamed CD patients (p=0.00097). This suggests that *SGPL1* is a Crohn's specific marker of inflammation and is differentially expressed between inflamed CD patients and controls.

TUSC3 is involved in the UPR, a pathway that has already been implicated in IBD. Data for *TUSC3* expression in ileal biopsies was not available, future work could focus on analysis expression in patient samples to confirm whether TUSC3 expression is important in IBD. Interestingly, *TUSC3* expression patterns suggest that *TLE1* functions as an activator, similarly to the situation with *RIOK1*.

The second part of this chapter aimed to address the effects of MDP mediated *NOD2* stimulation. However, this was not possible as NOD2 was excluded from the beadchip results as it did not meet the quality control criteria. Additionally, RT qPCR analysis did not confirm that *NOD2* stimulation was successful. The exact reasons underlying ineffective stimulation are not clear, they may be a result of a technical difficulties in mRNA quantification. Some difficulties have been experienced with the *NOD2* quantification in our laboratory due to low absolute expression levels, this was also observed in the beadchip data. MDP is a well characterised *NOD2* stimulant however other *NOD2* stimulants, such as TNFα could be used alongside MDP in future experiments. Additionally, *NOD2* overexpression constructs could be used. This would allow for confirmation of *NOD2* stimulation using multiple mechanisms and comparisons between activation of downstream signalling pathways. Recent work shows that MDP mediated stimulation of *NOD2* leads to dissociation of NOD2 from the chaperone protein, Hsp90 and subsequent proteasome mediated degradation of NOD2. This suggests that even though MDP stimulation upregulates *NOD2* expression, this may not directly correlate with NOD2 protein levels. Hence, *NOD2* mRNA and protein expression may not correspond and analysis of expression at both levels is essential to future work.

In these experiments the HEK 293 cell line was used, as previously discussed this cell line has been used extensively in IBD research. Research using this cell line primarily focuses on analysing the effects of *NOD2* stimulation as endogenous *NOD2* expression is low. In this chapter the effects of *NOD2* stimulation were analysed by comparing unstimulated and stimulated cells. In theory, the use of a cell line with very low *NOD2* expression would allow us to analyse the difference between unstimulated and

stimulated cells more easily as the differences would be expected to be more dramatic. However, the use of this cell line may have also prevented these experiments from detecting/effectively quantifying *NOD2* expression leading to *NOD2* probes on the expression chip failing to meet quality control criteria. Future work could focus on analysing *NOD2* expression in primary culture cells from the intestinal ileum, these cells have been shown to express *NOD2* at detectable levels, and furthermore, the results would provide a clearer insights into IBD pathogenesis.

It is important to note that in this study expression of all the genes discussed has been analysed at the mRNA level, this may not be a true representation of protein levels of the genes in question. It is thought that in human cell lines, mRNA expression can explain approximately 27% of the variance in protein concentration hence further work analysing protein levels is necessary to confirm these findings. Additionally, analysis of gene expression and localisation in different cell types of the ileum is necessary to fully understand the role of these genes in IBD.

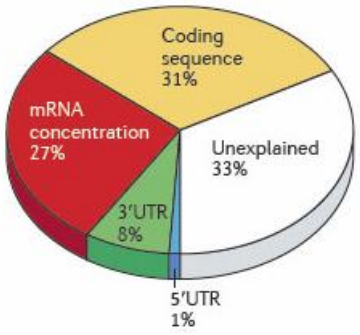


Figure 31 Contribution of mRNA concentration to variance in protein concentration in human medulloblastoma cell line

Pie chart showing factors influencing protein abundance adapted from Vogel *et al*, 2012. Data is based on the human medulloblastoma cell line, DAOY. mRNA expression is shown to account for approximately 27% of variance in protein expression and hence cannot be used as a definitive indicator of protein expression.

4 Investigating a potential *TLE1*/XBP1 interaction

4.1 Introduction

4.1.1 XBP1 and IBD

As discussed in the introduction, SNPs in the *XBP1* gene have been associated with IBD. Deep sequencing of the *XBP1* gene region revealed 4 variants that were only present in IBD patients, two of which are nonsynonymous SNPs that lead to hypomorphic forms of XBP1 (Kaser et al., 2008).

Cre recombinase based transgenic mouse models have shown that conditional deletion of *XBP1* in intestinal epithelial cells results in ER stress, characterised by increased splicing of *XBP1* mRNA and increased expression of the ER chaperone, Binding Immunoglobulin Protein (BiP). This ER stress results in reduced numbers of Paneth and Goblet cells. The few remaining Paneth cells are malformed, they lack the expanded ER seen in normal Paneth cells and have fewer granules which contain antimicrobial peptides such as α defensins. In summary, XBP1 null mice show increased ER stress, develop spontaneous enteritis and an exaggerated response to proinflammatory cytokines (IBD inducers) e.g. TNF α and flagellin in the epithelium. They also have an increased susceptibility to colitis as a result of Paneth cell absence and malformation (Kaser et al., 2008).

4.1.2 A potential XBP1 binding site in *TLE1*

As discussed in previous chapters, the rs6559629 SNP in *TLE1* is associated with IBD. One of the principal aims of this thesis is to investigate the pathological significance of this association.

Haploreg is a program developed by the Broad institute, it uses data from the 1000 genomes project to provide information about chromatin state, conservation and regulatory/transcription factor binding motifs alterations for a given SNP (Ward & Kellis, 2012a). It has been used to classify the functional significance of GWAS SNPs in many different contexts (Berndt et al., 2013; Bønnelykke et al., 2014; Chung et al., 2013; Cui et al., 2013; Hinds et al., 2013; Rhie et al., 2013).

Using the Haploreg program we identified that rs6559629 interrupts two potential XBP1 binding sites in *TLE1* (Figure 32).Figure 33 shows the XBP1 core binding motifs, the rs6559629 site interrupts the ACGT and UPRE core binding motifs. It is important to note that the core motifs of these sequences only differ by a single base pair. The result of this SNP is to change a CpG site. This site is created in the presence of the G allele of the SNP and destroyed by the A allele.

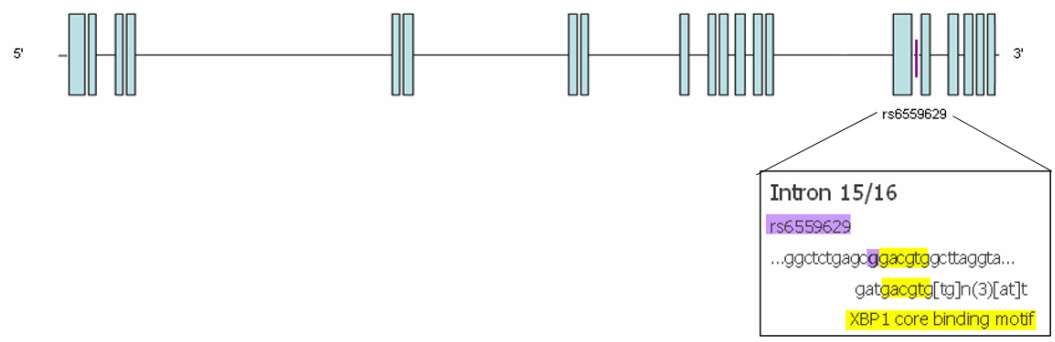


Figure 32 Schematic diagram showing XBP1 binding site in *TLE1*

Schematic diagram showing *TLE1* gene, exons are represented by blue boxes, introns are represented by black line joining exons. The rs6559629 SNP is shown as a purple line located in intron 15/16. The box shows the XBP1 core binding motif in yellow and the rs6559629 in purple.

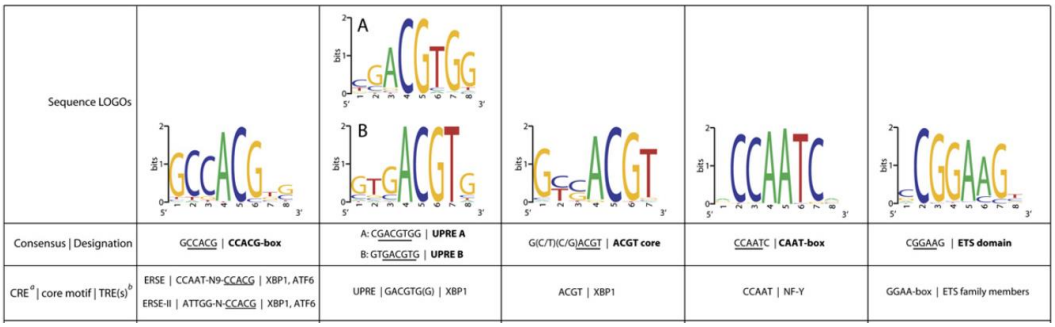


Figure 33 XBP1 core binding motifs

Figure adapted from (Acosta-Alvear et al., 2007). Table showing names and sequences of core XBP1 binding motifs. These motifs were identified as over represented in XBP1 targets in a chip on chip experiment conducted by Acosta-Alevear et al. Both ACGT and UPRE binding sites are interrupted by the rs6559629 SNP in *TLE1*.

4.1.3The Endoplasmic reticulum (ER) and the Unfolded Protein Response (UPR)

The rough endoplasmic reticulum (ER) is the eukaryotic organelle responsible for the synthesis, folding and post translational modification of proteins. Following synthesis and folding in the ER, proteins are transported to the Golgi apparatus in transition vesicles. Proteins are then secreted in vesicles to their final destination. Under normal conditions, any unfolded/incorrectly folded proteins are targeted for ER associated degradation (ERAD) in the ER and then transported to the cytoplasm where they are degraded by proteasomes. ER chaperones responsible for correct protein folding as well as ERAD chaperones and enzymes are constitutively expressed in order to process nascent proteins. When the ERAD cannot cope with all the unfolded/misfolded proteins, they accumulate in the ER, this situation is defined as ER stress. Prolonged ER stress initiates apoptosis. A number of factors including viral infection, glucose depletion and alterations in ion concentration have been shown to cause ER stress. Glucose depletion is thought to alter N linked glycosylation of proteins in the ER and activation of the PERK pathway (De La Cadena, Hernández-Fonseca, Camacho-Arroyo, & Massieu, 2014). Viral infection and replication increases the burden on ER protein synthesis chaperones leading to ER stress and apoptosis (Tardif, Mori, &

Siddiqui, 2002). The ER is a calcium rich environment and many ER chaperones are Ca^{2+} dependant, alterations in Ca^{2+} concentration lead to protein misfolding and therefore UPR activation (reviewed in (Michalak, Parker, & Opas, 2002) . In response to ER stress, the unfolded protein response (UPR) is activated to restore homeostasis (reviewed in (Ma & Hendershot, 2001). The UPR initiates signalling pathways in order to resolve this stress. These pathways function to inhibit protein translation, produce more ER chaperones and ERAD components, increase production of secretory cells and initiate apoptosis in damaged cells.

The UPR signalling pathways

There are three UPR signalling pathways: the Ire1α pathway, the ATF6α pathway and the PERK signalling pathway. A summary of these three pathways is shown in Figure 29.

The presence of misfolded proteins in the ER is detected by ER sensors and leads to activation of the three UPR signalling pathways. All three pathways are activated by dissociation from the ER chaperone, Binding immunoglobulin protein (BiP). When PERK is released from BiP it oligomerises and undergoes phosphorylation becoming an active serine/threonine kinase. PERK initiates downstream signalling pathways leading to activation of transcription factors such as C/EBP homologous protein (CHOP), Activating transcription factor 6 (ATF6) and Eukaryotic initiation factor 2α (E1F2α). These transcription factors activate transcription of antioxidative, ER chaperone and apoptosis genes.

In the ATF6α pathway, dissociation from BiP allows ATF6α to translocate to the Golgi apparatus and be cleaved by the site 1 and site 2 proteases (S1P and S2P). ATF6α moves from the cytosol into the nucleus where it initiates transcription of genes involved in protein folding and ER associated degradation (ERAD).

The most evolutionarily conserved of the three UPR signalling pathways is initiated by inositol requiring transmembrane kinase endoribonuclease 1 (IRE1). IRE1 has two isoforms; IRE1α which is ubiquitously expressed and IRE1β which is expressed in the intestinal epithelium. Under normal conditions IRE1 is inactive as it is bound by the chaperone BiP. IRE1 is a sensor/transducer of ER stress and is activated in the presence of misfolded/unfolded proteins and released from BiP. IRE1α oligomerises and undergoes autophosphorylation, this activated form of IRE1α is an endonuclease and kinase. Activated IRE1α cleaves 26nt from the mRNA transcript of the inactive, unspliced form of *XBP1* (Calfon et al., 2002). Unlike the unspliced form of XBP1 (XBP1us), the spliced form of XBP1 (XBP1s) is an active transcription factor that binds to the promoter regions of target genes. XBP1 has 545 known target genes, of which a "core" of 95 genes are found in more than one cell type (Acosta-Alvear et al., 2007). These target genes were identified in both ER stressed and unstressed cells in an XBP1 chip on chip experiment using a mouse promoter array (Acosta-Alvear et al., 2007). These promoter arrays covered DNA up to 800bp upstream and 300bp downstream of approximately 14,000 known mouse genes. A large proportion of XBP1 target genes were shown to be UPR related, they are involved in protein folding, trafficking, secretion and synthesis. However, XBP1(s) targets are not limited to UPR related genes, XBP1 (s) also activates transcription of genes involved in DNA replication and repair, cell growth, differentiation and glucose homeostasis

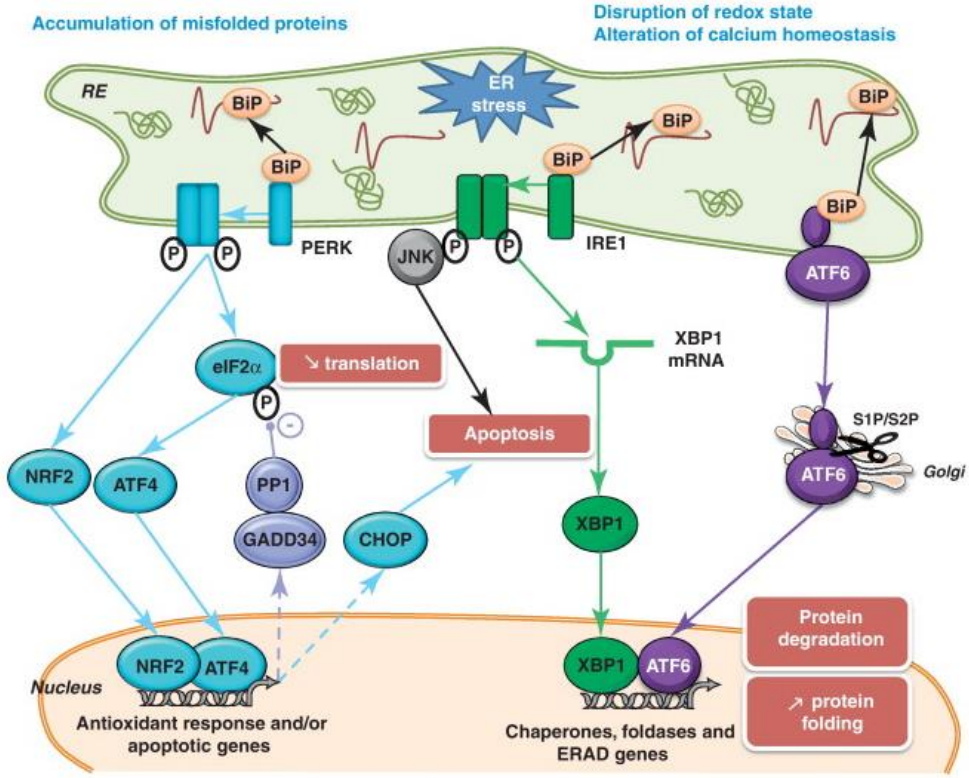


Figure 34 Summary of UPR signalling pathways

Summary of the PERK, Ire1α and ATF6 UPR signalling pathways. Binding Immunoglobulin protein (BiP) is the chaperone protein that in response to ER stress dissociates from the sensors of all three pathways.

In the PERK pathway dissociation of BiP from PERK leads to phosphorylation and oligomerisation of PERK, converting it to an active serine/threonine kinase. In response to activation, PERK phosphorylates E1F2α which stops translational initiation, thereby decreasing the protein folding load on the ER. Downstream signalling also leads to activation of the transcription factors: NRF2 (Nuclear factor 2), ATF6 (Activating transcription factor 6) and CHOP (C/EBP homologous protein). These initiate transcription of antioxidative, ER chaperone and apoptosis genes respectively.

IRE1α is activated similarly to PERK, activated IRE1α is an endonuclease which cleaves 26nt from the unspliced *XBP1* mRNA transcript. The cleaved *XBP1* mRNA transcript is translated into an active transcription factor, initiating transcription of genes involved in protein synthesis, degradation, folding and ER expansion.

The ATF6α protein translocates to the Golgi apparatus upon activation and is cleaved by the site 1 and site 2 proteases (S1P and S2P). It then moves from the cytosol into the nucleus where it initiates transcription of genes involved in protein folding and ER associated degradation (ERAD).

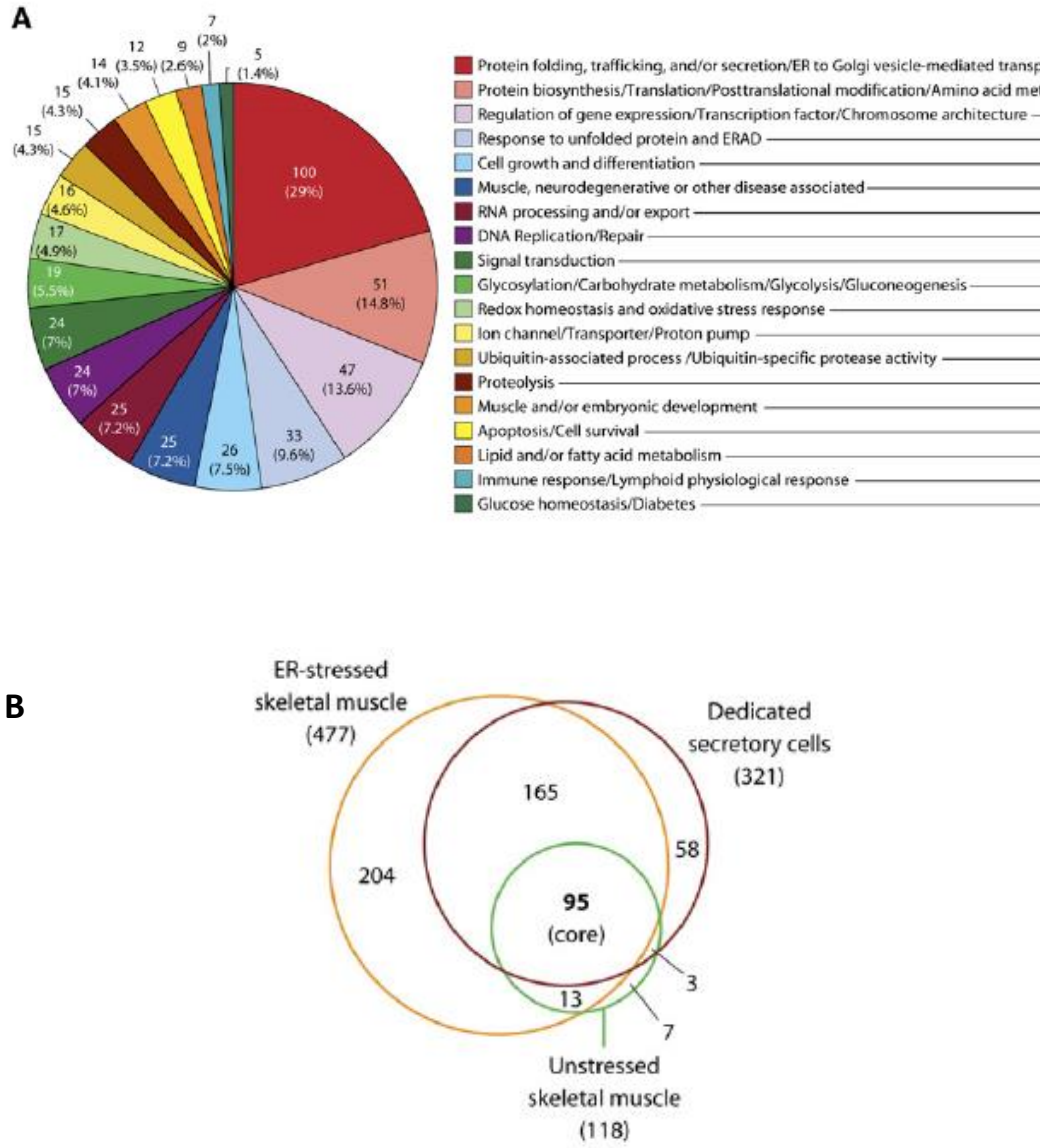


Figure 35 Summary of XBP1 target genes

Figure adapted from Acosta et al, 2007. A) Pie chart summarising number of XBP1 target genes in each Gene Ontology (GO) category. The percentage represents the proportion of genes in each GO category out of a total of 545 target genes. B) Venn diagram showing the number of XBP1 target genes that are specific to each of the three cell types used in this study as well as the 95 core genes expressed in all three cell lines (Acosta-Alvear et al., 2007).

4.1.4 ER stress, the UPR and IBD

ER stress and the UPR dramatically impact the function of the intestinal epithelium; Paneth and Goblet cells are highly secretory cells and therefore need to produce correctly folded proteins. A defective UPR has a dramatic impact on the secretory capacity of these cells.

ER stress and activation of the UPR has been shown in both CD and UC. For example, in CD increased expression of BiP and XBP1(s) has been shown by immunohistochemistry in intestinal biopsies (Deuring et al., 2011; Kaser et al., 2008). IRE1β has also been implicated in IBD; IRE1β-/- mice show increased BiP expression and sensitivity to DSS induced colitis (Iqbal et al., 2009). Associations between other UPR related genes such as *AGR2* and *ORMDL3* are discussed in Chapter1.

4.1.5 Chromatin Immunoprecipitation

Chromatin Immunoprecipitation (ChIP) is a technique used to determine whether a protein binds to a specific region of DNA. Briefly it involves, cross linking protein to DNA, sonication of DNA, immunoprecipitation with an antibody to the protein of interest, purification of DNA and PCR using primers for the region of interest. ChIP has previously been used to identify XBP1 target genes such as *Xbp1* and *Atf6* in mouse embryonic fibroblasts (Acosta-Alvear et al., 2007).

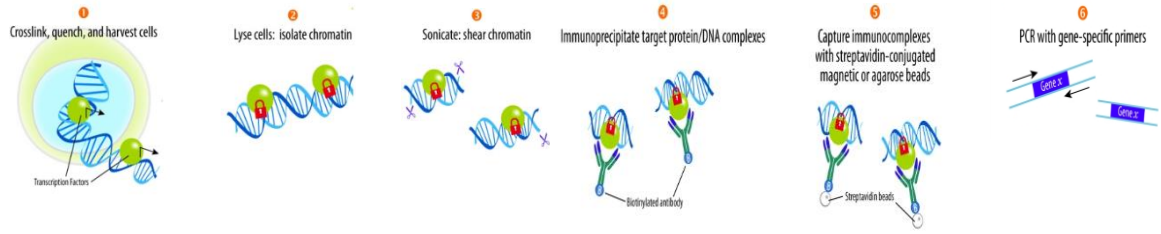


Figure 36 Chromatin Immunoprecipitation

Schematic diagram showing the key stages of chromatin immunoprecipitation.1) Cells are treated with a fixing agent such as paraformaldehyde to cross link the protein to the DNA. 2) Cells are lysed, DNA/protein complexes are harvested. 3) DNA and bound proteins are sonicated (usually to approximately 500bp). 4) and 5) Antibodies to the protein of interest (in this case XBP1) are used to isolate DNA bound to the protein. Following immunoprecipitation of DNA bound to the protein of interest, the DNA/protein cross links are reversed. DNA is isolated from the protein, this is the DNA that the protein of interest binds to. 6) Primers for the site of interest (in this case the rs6559629 site) are used in RT qPCR to determine whether the protein of interest binds to the site.

4.2 Aims

One of the principal aims of this thesis is to elucidate the mechanism by which rs6559629, a SNP associated with ileal CD, may be contributing to disease pathogenesis. This SNP has been shown to interrupt a potential XBP1 binding site. XBP1 is involved in the unfolded protein response and is a known IBD susceptibility gene (Kaser et al., 2008). The aim of this chapter was to determine whether XBP1 binds to this predicted binding site. This aim can be broken down into several stages:

1. Design a ChIP assay for XBP
    a. Optimise DNA sonication
    b. Identify and characterise a correctly functioning antibody
    c. Characterise positive and negative controls
    d. Design and validate primers for the rs6559629 binding site

2. Determine whether XBP1 binds to the predicted rs6559629 binding site in *TLE1*.

4.3 Results

4.3.1 Sequencing rs6559629 in cell lines

Primers were designed for the rs6559629 region in *TLE1* in order to determine the rs6559629 genotype in HCT116, HT29, SW480 and HEK293 cells (

Table 13). Cells were plated and harvested the following day at 80% confluency. DNA was extracted using a Qiagen Dneasy kit and amplified by PCR using primers for the rs6559629 region (see Appendix for details) and standard PCR protocols (seen Chapter 2). Sanger sequencing technology was used to sequence the resultant PCR products. Sequence was analysed using Geneious r7 software. The three cell lines used in this experiment were HEK 293, HCT116 and SW480 cells each of which have a different rs6559629 genotype. The use of cell lines with all three different genotypes allows us to determine whether XBP1 binds differentially to different alleles of the rs6559629 SNP. All three cell lines have been used extensively in IBD research.

Cell Line	Description	Rs6559629 genotype
HEK293	Human Embryonic kidney	GG
HCT116	Human colonic carcinoma	GA
SW480	Human colonic carcinoma	AA

Table 13 Cell lines and rs6559629 genotype

Description and rs6559629 genotype of the three cell lines used in this experiment: HEK293, HCT116 and SW480.

4.3.2 Optimisation of sonication conditions

The first stage of ChIP involves cross linking DNA to protein and sonicating the DNA/protein complexes (Figure 36). The optimal size of DNA fragments to be used in ChIP experiments is between 200-800bp, for this experiment an average DNA fragment size of 500bp was used.

Cells were plated in 10cm plates, the following day (80% confluency) DNA bound proteins were cross linked to DNA by incubating cells in 10mls DMEM supplemented with 1% formaldehyde (10 mins, 37˚C, 5% CO_2). Following crosslinking, the formaldehyde solution was aspirated and cells were incubated with 10mls of DMEM supplemented with 0.125M glycine (Sigma) (10 mins, room temperature on a rocking platform). Cells were washed twice in ice cold PBS and scraped into 1ml PBS. Samples were centrifuged (5 mins, 5000 RPM, 4˚C) and PBS was removed. Each cell pellet was lysed in 200µl fresh ChIP lysis buffer (see Chapter 2) on ice for 10 mins. Samples was sonicated using a Bioruptor standard water bath sonicator (Diagenode) for 10, 15 and 20 cycles of sonication (30 seconds on, 30 seconds off at 4˚C). Following sonication, samples were centrifuged (15 mins, 13000 RPM, 4˚C). In order to determine the size of sonication fragments, cross links were reversed at this stage, for the full ChIP experiment following sonication there was no reversal of cross links and samples were used for XBP1 immunoprecipitation. To reverse cross links and check DNA fragment size all samples were phenol chloroform treated. This involved: adding 5µl 4M sodium chloride, incubating at 65˚C overnight, adding a mixture of 50 µl phenol, 48 µl chloroform and 2 µl isoamyl alcohol vortexing (10 mins) and centrifuging (5 mins, 13000 RPM, 4˚C). The upper phase of the resulting mixture contains DNA, 10 µl of 2M sodium acetate and 220 µl ice cold ethanol was added to 100 µl of the upper phase. Samples were centrifuged (5 mins, 13000 RPM, 4 ˚C) and the supernatant was discarded. The resulting DNA pellet was washed in 1ml ethanol, centrifuged (5 mins, 10000 RPM, 4 ˚C) and air dried. Each DNA pellet was resuspended in 30 µl TE buffer. The sample was vortexed and incubated at -80 ˚C (1 hour). Each sample (10 µl) was run on a 1.5% agarose gel as shown in Figure 37. The optimal average sonication size of 500bp was achieved after 20 cycles of sonication (Figure 37).

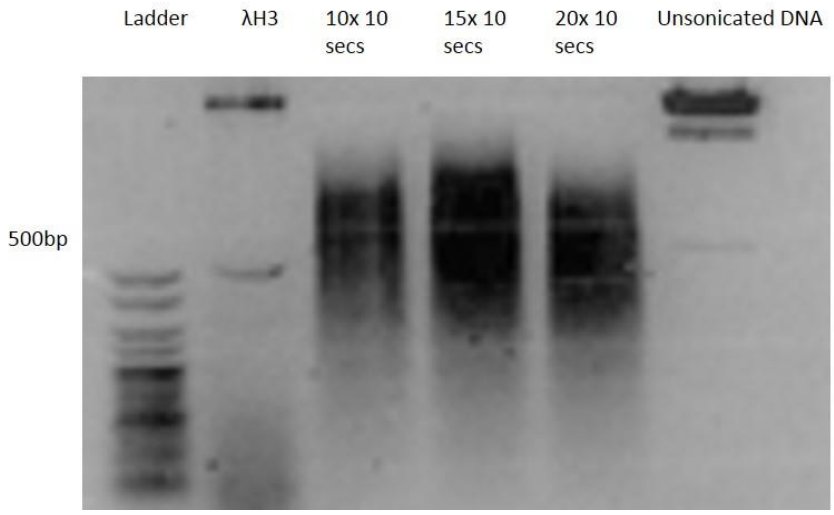


Figure 37 Optimisation of sonication conditions

Image of 1.5% agarose gel showing Hyperladder V (Invitrogen), DNA digested with Lamda HindIII, sonicated DNA (10, 15 and 20 sonication cycles of 10 seconds off, 10 seconds on) and unsonicated DNA. The 500bp mark is shown on the left hand side.

4.3.3 Immunoprecipitation of XBP1 (spliced and unspliced)

Following cross linking of protein to DNA, cell lysis and sonication the fourth stage of ChIP involves immunoprecipitation (IP) of DNA bound to the protein of interest (XBP1) (Figure 36). In order to determine whether the XBP1 antibody recognised the XBP1(s) and XBP1 (us) proteins, cells were transfected with XBP1 over expression constructs and incubated with anti XBP1 antibody on a western blot. Plasmids containing either XBP1 spliced (pXBP1SP) or XBP1 unspliced (pXBP1US) were a kind gift from Professor Arthur Kaser (Kaser et al., 2008).

Cells were plated and transfected the following day with 10µg of empty vector, pXBP1SP or pXBP1US according to standard transfection protocols and incubated for 24 hours (37˚C, 5% CO_2). Cells were sonicated, cross linked and lysed as described above. Prior to setting up the IP, 30µl of cell lysate was kept as input sample. Protein G agarose beads (Roche) used in this experiment were washed in ChIP dilution buffer (see Chapter 2) and centrifuged (5mins, 10000RPM, 4 ˚C) prior to use. Before

immunoprecipitation, 100 µl of protein lysate was incubated with 50µl protein G agarose beads, 1 µl rabbit IgG (Sigma), 2 µl salmon sperm DNA (Roche) and 20 µl purified water. Samples were pre cleared for 6 hours on a rotating wheel at 4 ˚C to remove any components of the reaction mixture that may bind to the agarose beads nonspecifically. Samples were centrifuged (5 mins, 2000RPM, 4 ˚C) and the supernatant transferred to a new Eppendorf tube. Following preclearing, DNA bound to XBP1 was immunoprecipitated overnight. The IP reaction contained: 5µl anti XBP1 antibody (M186, Santa Cruz), 30µl protein G agarose beads, 100µl cell lysate and 965µl purified distilled water. For the purposes of characterising the M186 XBP1 antibody, 30 µl loading dye was added to 30 µl of each sample and the samples was heated for 10 mins at 45 ˚C. Samples were run on a 10% polyacrylamide gel according to standard protocols and blotted with M186 anti XBP1 antibody at a concentration of 1/1000, overnight. Figure 38 shows that both the unspliced (55KDa) and spliced forms of XBP1 (35KDa) are detected on a western blot in HEK293 cells transfected with the relevant constructs. In cells transfected with the unspliced XBP1 construct, the spliced construct can also be detected at 55KDa. These results confirmed that the M186 XBP1 antibody could detect the XBP1 protein on a western blot. In addition, the lack of bands in the empty vector transfection shows that XBP1 does not detect any nonspecific proteins on a western blot. The M186 XBP1 antibody has previously been shown to bind to XBP1 and known XBP1 targets (Acosta-Alvear et al., 2007) .

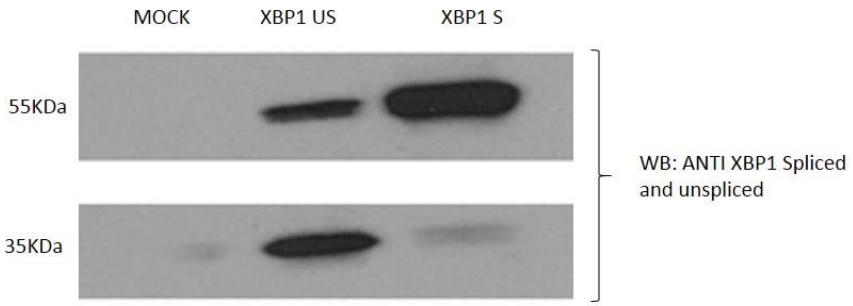


Figure 38 Confirming immunoprecipitation of XBP1 protein

Image showing western blot using 10% agarose gel. Top image was blotted with anti XBP1 (S) and (US) antibody (M186, Santa Cruz).Three samples were run on each gel (left to right): Mock (empty vector) transfected HEK 293 cells, HEK293 cells transfected with pXBP1SP and HEK293 cells transfected with pXBP1US. Bands at 55KDa show XBP1 (US), bands are 35KDa show XBP1(S).

4.3.4 *XBP1* promoter region as a positive control

XBP1 has previously been shown to bind to the *XBP1* promoter region as part of a positive feedback loop (Acosta-Alvear et al., 2007). In order to confirm that the M186 XBP1 antibody was binding to XBP1 and the ChIP assay was working, the *XBP1* promoter region was used a positive control for this ChIP experiment. Although the *XBP1* promoter region has previously been used as a positive control for ChIP experiments, the exact co-ordinates of the region where XBP1 binds to the *XBP1* promoter were not available. In order to identify the region in question the UCSC genome browser was used to analyse histone acetylation of the promoter region (Figure 39B). Histone acetylation is a marker for open chromatin which indicates transcriptionally active regions of DNA, as would be expected of a promoter region. Primers for regions between 200 and 250 bp long were designed for the *XBP1* promoter region, where histone acetylation appeared most dense (29,196,500-29,198,500).

To determine whether XBP1 was binding to this site, the entire ChIP experiment was conducted on HEK293 cells transfected with an empty vector, pXBP1S or pXBP1US

construct. Samples were lysed, sonicated and immunoprecipitated overnight as described above. Following immunoprecipitation, samples were centrifuged (5 mins, 2000RPM, 4 ˚C) and the supernatant was discarded. Beads were washed in 1ml of the following buffers: TSE, TSEII, buffer III and twice in TE. Each wash lasted 5 mins on a rotating wheel, in between washes beads were centrifuged (5 mins, 2000RPM, 4 ˚C) and the supernatant was removed. After the final wash, the DNA was eluted from the protein G agarose beads in 250µl freshly made 1% sodium bicarbonate solution for 15 mins on a rotating wheel at 4˚C. Supernatant was removed (kept in a separate Eppendorf) and elution was repeated in 200µl 1% sodium bicarbonate solution (15 mins, 4 ˚C). The solution from both elution's (450µl) was collected in one Eppendorf and 25µl of 4M sodium chloride was added (1.75µl was added to input samples). Samples were heat treated at 65 ˚C overnight. See Chapter 2 for further details on buffers used. The following day, 30µl of each sample was purified using a QIAquick PCR purification kit (Qiagen). DNA was quantified by RT qPCR using standard protocols and undiluted DNA samples.

Results from this XBP1 ChIP experiment of the *XBP1* promoter show that XBP1 binds to the *XBP1* promoter region and that this region can be used as a positive control (**Error! Reference source not found.**). **Error! Reference source not found.** shows RT qPCR analysis of HEK 293 cells transfected with empty vector or pXBP1SP for the 29,198,500 to 29,196500 region of the *XBP1* promoter. Fold enrichment was calculated by calculating concentration values ($2^{(-CT)}$) and normalising all concentrations to the empty vector transfection. Each bar represents average fold enrichment across three replicates. A fold enrichment of between 6.8 and 7.5 (± SEM 0.03-0.21) fold is seen between co-ordinates 29,197,700 and 29,196,900.

Two positive control regions (region 1 = 29,197,500-700, region 2= 29,196,900-7100) were used in the final ChIP experiment in HEK 293, HCT 116 and SW 480 cells. **Error! Reference source not found.** shows the results from this experiment, all three cell lines were transfected with either an empty vector, pXBP1S or pXBP1US construct. Bars show average fold enrichment normalised to the empty vector transfection. Results shown are representative of three biological replicates. In HEK

293 cells transfected with a pXBP1US there was an average fold enrichment of 1.41 (± SEM 0.29) in region 1 and 1.73 (± SEM 0.47) in region 2. In HEK 293 cells transfected with pXBP1SP there was an average fold enrichment of 7.57 (± SEM 0.75) in region 1 and 6.24 (± SEM 2.95) in region 2. . In HCT 116 cells transfected with pXBP1US there was an average fold enrichment of 1.94 (± SEM 0.45) in region 1 and 1.65 (± SEM 0.33) in region 2. In HCT 116 cells transfected with pXBP1SP an average fold enrichment of 6.14 (± SEM 1.08) for region 1 and 8.56 (± SEM 0.25) for region 2 was observed. In SW 480 cells transfected with pXBP1US there was an average fold enrichment of 1.16 (± SEM 0.11) in region 1 and 1.80 (± SEM 0.40) in region 2. In SW 480 cells transfected with pXBP1SP there was an average fold enrichment of 7.85 (± SEM 1.41) for region 1 and 11.11 (± SEM 2.13).

Overall, all three cell lines, when transfected with unspliced XBP1, showed a fold enrichment between 1.16 and 1.94 (± SEM 0.11-0.45) for region 1 and between 0.85 and 1.80 (± SEM 0.33-0.47) for region 2. On average, all three cell lines, when transfected with spliced XBP1, showed a fold enrichment between 6.14 and 7.85 (± SEM 0.75-1.41) for region 1 and between 6.24 and 11.11 (± SEM 0.25-2.95) for region 2.

4.3.5 Negative controls for XBP1 ChIP

In order to show that the XBP1 antibody was binding specifically to XBP1 targets, negative control regions were used in the final ChIP experiment. Two regions were used as negative controls (region 3 = 29,197,700-900, region 4= 29,196,700-900) in HEK293, HCT116 and SW480 cells. As above, bars show average fold enrichment normalised to the empty vector transfection and are representative of three biological replicates. In HEK 293 cells transfected with pXBP1US there was an average fold enrichment of 0.84 (± SEM 0.26) in region 3 and 1.73 (± SEM 0.41) in region 4. In HEK 293 cells transfected with pXBP1SP there was an average fold enrichment of 0.69 (± SEM 0.26) in region 3 and 1.90 (± SEM 0.85) in region 4. In HEK 293 cells transfected with pXBP1US there was an average fold enrichment of 0.84 (± SEM 0.26) in region 3 and 1.73 (± SEM 0.41) in region 4. In HCT 116 cells transfected with

pXBP1US there was an average fold enrichment of 2.05 (± SEM 1.31) in region 3 and 0.98 (± SEM 0.20) in region 4. In HCT 116 cells transfected with pXBP1SP there was an average fold enrichment of 2.00 (± SEM 0.47) in region 3 and 0.85 (± SEM 0.10) in region 4. In SW 480 cells transfected with pXBP1US there was an average fold enrichment of 0.91 (± SEM 0.17) in region 3 and 0.94 (± SEM 0.17) in region 4. In HCT 116 cells transfected with pXBP1SP there was an average fold enrichment of 1.06 (± SEM 0.55) in region 3 and 1.46 (± SEM 0.27) in region 4.

Overall, all three cell lines, when transfected with unspliced XBP1, showed a fold enrichment between 0.84 and 1.31 (± SEM 0.17-1.31) for region 3 and between 0.94 and 1.73 (± SEM 0.16-0.41) for region 4. On average, all three cell lines, when transfected with spliced XBP1, showed a fold enrichment between 0.69 and 2.00 (± SEM 0.26-0.54) for region 3 and between 0.85 and 1.90 (± SEM 0.10-0.85) for region4,

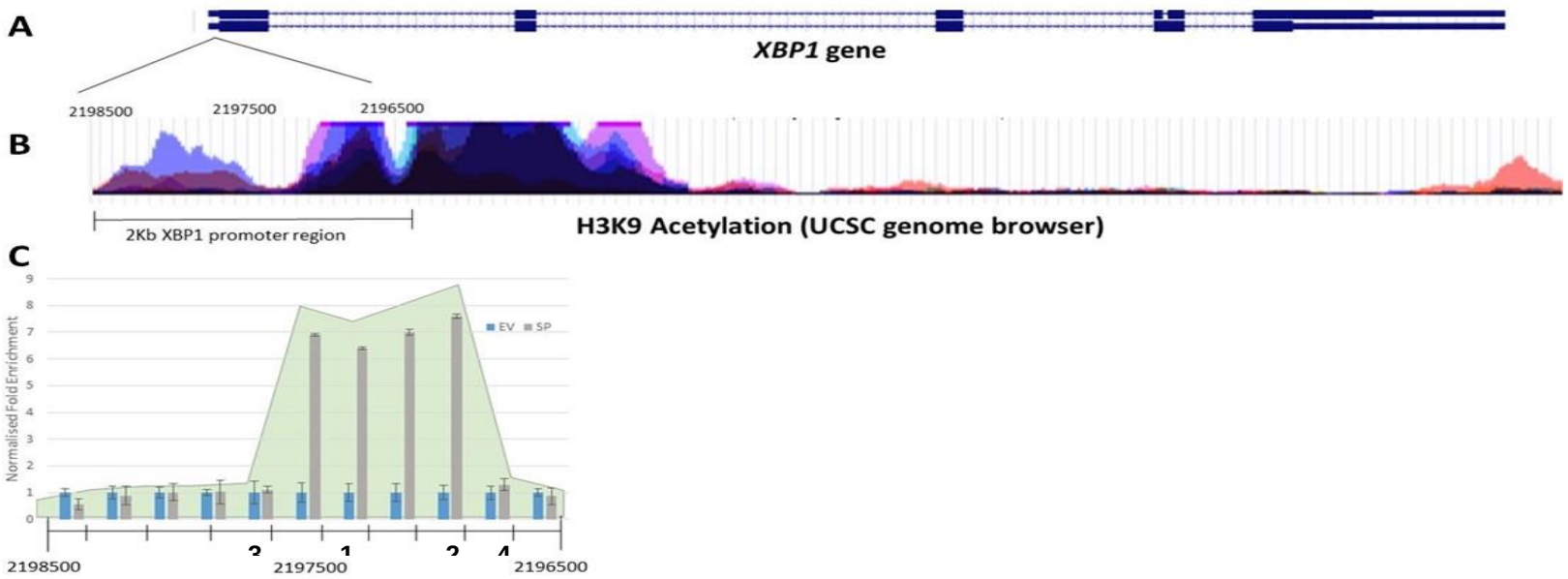


Figure 39 Identification and characterisation of XBP1 binding site in *XBP1* promoter region

Figure showing the *XBP1* gene (A), corresponding H3K9 acetylation track (B) (UCSC) and graph results from an XBP1 ChIP experiment on this region (C). This region has been used as a positive control in XBP1 ChIP experiments (Acosta-Alvear et al., 2007). Bars represent average fold enrichment normalised to empty vector (EV) transfection. Blue bars show fold enrichment from HEK 293 cells transfected with an empty vector (EV)j, grey bars show fold enrichment for HEK 293 cells transfected with spliced XBP1 construct. Error bars show the standard error of the mean across three biological replicates. Numbers 1, 2, 3 and 4 correspond to positive and negative control regions used in final experiments.

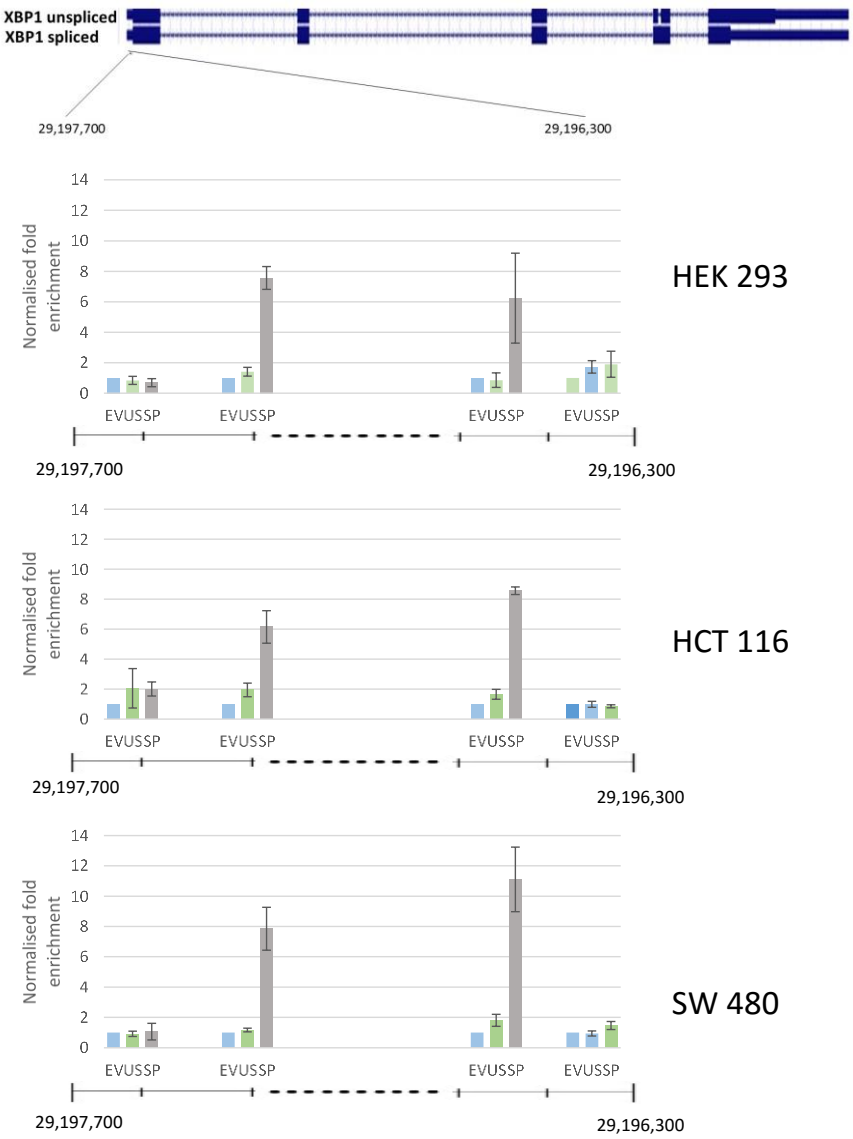


Figure 40 Positive and negative controls for XBP1 ChIP experiment

XBP1 ChIP results for the XBP1 promoter region (19,196,300-29,197,700) are shown as average fold enrichment, normalised to empty vector (EV) transfections. Blue bars (EV) represent cells transfected with an empty vector, green bars (US) represent cells with pXBP1US construct and grey bars pXBP1SP. Each bar shows average fold enrichment across three biological replicates. Error bars represent the standard error of the mean.

4.3.6 XBP1 does not bind to the rs6559629 site

Error! Reference source not found. shows positive and negative controls for this XBP1 ChIP experiment for the three cell lines to be used in this experiment. These same samples were used to determine whether XBP1 binds to its predicted binding site, the rs6559629 site in TLE1. Each of these cell lines has a different rs6559629 genotype to allow quantification of differential binding (

Table 13). Figure 8 shows average fold enrichment of the 1Kb region surrounding the rs6559629 SNP. This 1Kb region was chosen as results from analysis of the *XBP1* promoter region showed that XBP1 bound a region approximately 500bp away from the core binding motif (**Error! Reference source not found.**).

Figure 41 shows the XBP1 ChIP results for the rs6552629 region (84,206,860-207860) in HEK 293, HCT 116 and SW480 cells transfected with either empty vector, pXBP1US or pXBP1SP constructs. Bars show the average fold enrichment normalised to empty vector transfections, as described earlier. In HEK 293 cells (rs6559629 genotype, GG) transfected with pXBP1US, there was an average fold enrichment between 0.98 and 1.31 (± SEM 0.20-0.53) across the rs6559629 region. In HEK 293 cells transfected with pXBP1SP there was an average fold enrichment between 0.87 and 1.44 (± SEM 0.06-0.56). In HCT 116 cells (rs6559629 genotype, GA) transfected with pXBP1US there was an average fold enrichment between 0.75 and 1.12 (± SEM 0.15-0.43). In HCT 116 cells transfected with pXBP1SP there was an average fold enrichment between 0.87 and 1.33 (± SEM 0.11-0.33) across the rs6559629 region. In SW480 cells (rs6559629 genotype, AA) transfected with an pXBP1US there was an average fold enrichment between 1.11 and 2.06 (± SEM 0.33-0.46) .In SW480 cells transfected with pXBP1SP there was an average fold enrichment between 0.73 and 1.41 (± SEM 0.04-0.66) across the rs6559629 region.

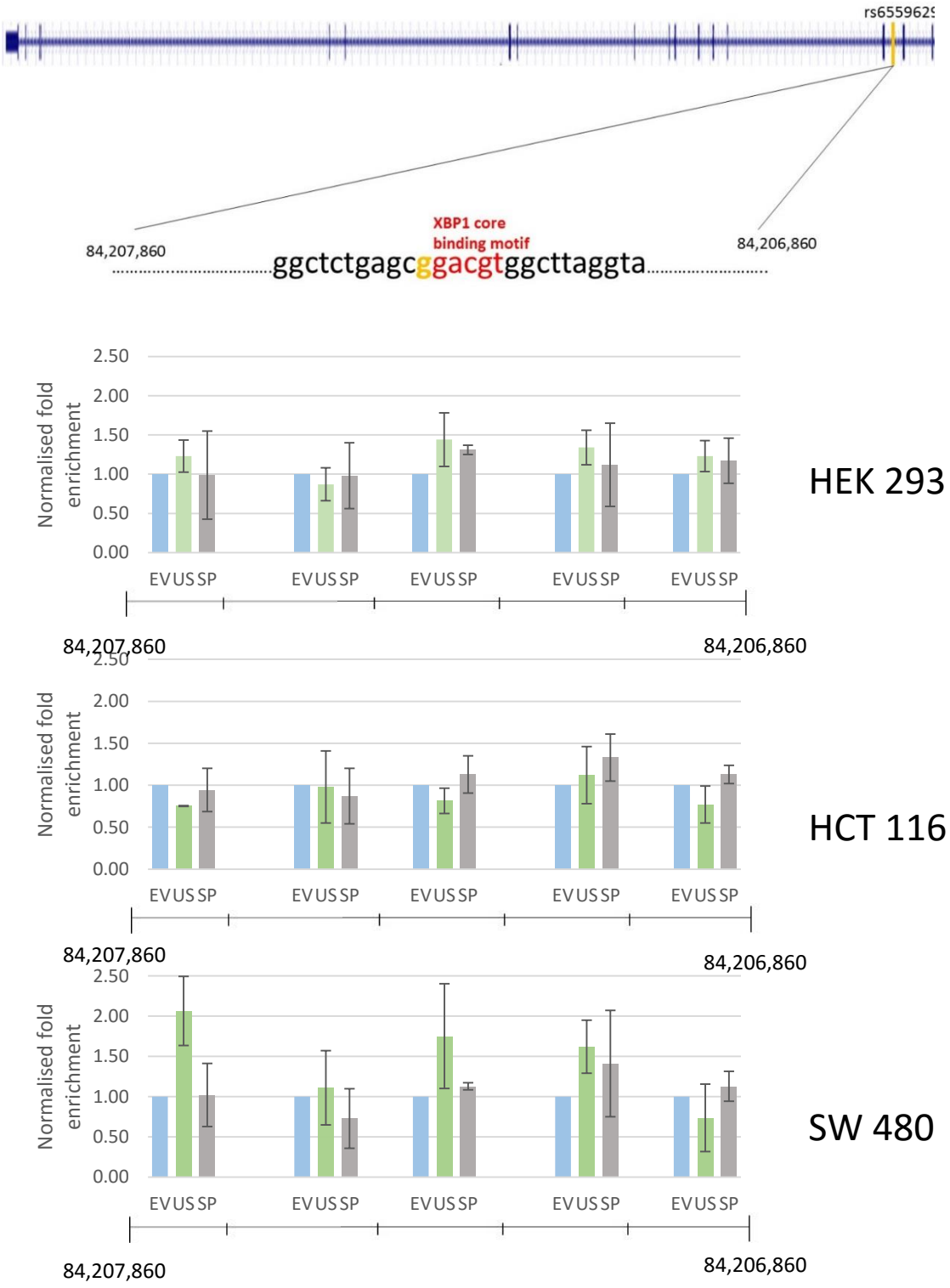


Figure 41 XBP1 does not bind to the rs6559629 site in HEK293, HCT116 and SW480 cells

XBP1 ChIP results for the rs6559629 region (84,207,860-84,206,860) are shown as average fold enrichment, normalised to empty vector (EV) transfections. Blue bars (EV) represent cells transfected with an empty vector, green bars (US) represent cells with pXBP1US construct and grey bars pXBP1SP. Each bar shows average fold enrichment across three biological replicates. Error bars represent the standard error of the mean.

4.4 Discussion

The aim of this chapter was to determine whether XBP1 binds to a predicted binding site in *TLE1*. This predicted binding site was of particular interest as it is interrupted by the rs6559629 SNP which is associated with Crohn's disease (Figure 32) (Nimmo et al., 2011).

This was achieved by designing a ChIP assay using an anti XBP1 antibody and RT qPCR to quantify DNA enrichment. Three cell lines were used in these experiments to quantify differential binding of XBP1 according to rs6559629 genotype (

Table 13). The *XBP1* promoter region was shown to be a positive control for this assay in all three cell lines. Adjacent negative control regions were also used. Both unspliced and spliced forms of XBP1 were over expressed in each cell line as endogenous levels of XBP1 are low. The results described show that XBP1 does not bind to the rs6559629 site in *TLE1* in any of the cell lines used.

The rs6559629 site is not in a known promoter region, the nearest transcriptional start site (TSS) is the *TLE1* TSS which is approximately 10Kb downstream of the XBP1 recognition motif, in intron 15/16 of *TLE1*. Many studies have focused on the binding of XBP1 to promoter regions for example XBP1 has been shown to bind to promoter regions of GPR43, CHOP and BECLIN1 using ChIP assays (Ang, Er, & Ding, 2015; Margariti et al., 2013; Shao et al., 2015). Although the potential binding site interrupted by rs6559629 was not close to a TSS it was targeted by this study as it had sequence overlap for two core XBP1 binding motifs. Additionally although Acosta et al suggest that many a majority of XBP1 binding motifs occur at transcriptional start sites, their work used proximal promoter arrays and hence a long range transcriptional activation mechanism for XBP1 cannot be ruled out. Many transcriptional regulators have been shown to have long range effects in addition to short range ones. For example, the transcriptional activator, estrogen receptor (ER) has only 22% of its targets near TSS, the remaining targets are long range targets, some of which are up to 100Kb away from a TSS. Furthermore these distal ER binding sites are functional and

activate transcription of downstream genes (Carroll et al., 2006). Although according to this work XBP1 does not bind the rs6559629 site, it would be of future interest to explore long range effects of XBP1 to further elucidate its role in IBD pathogenesis.

In this work, binding of both the spliced and unspliced forms of XBP1 were analysed. Recent work has shown that the XBP1 (us) is a functional interacting protein. It has been shown to interact with the autophagy protein, FoxO1 by co immunoprecipitation and co- localisation was shown using immunocytochemistry. It was also suggested that XBP1(us) facilitates degradation of FoxO1 by the proteasome (Zhao et al., 2013). Although the work presented suggests XBP1 (us) does bind to the same positive control regions as XBP1 spliced, it is possible that XBP1 (us) has different targets and mechanisms of action to XBP1 (s). However, it is important to note, that over expression of XBP1 (us) induces ER stress and expression of XBP1(s) and therefore the *XBP1* promoter region may not have been a positive control for XBP1 (us). Therefore it could be the case that this ChIP assay did not accurately detect the XBP1 (us) targets. The use of high throughput technologies such as chip on chip to identify XBP1 (us) targets would be of interest, to identify whether XBP1 (us) binds to the rs6559629 site as well as to compare XBP1 (s) and XBP1 (us) functionality.

The rs6559629 SNP interrupts a CpG site and the methylation status of this CpG site may influence XBP1 binding. As part of this project, the methylation status of this site was investigated in these cell lines however it was not possible to design an appropriate pyrosequencing assay following four attempts with various primers. Other work in our laboratory has analysed global methylation in IBD patients vs controls, which included four sites in *TLE1*, these sites did not include the rs6559629 site and were not differentially methylated between patients and controls (Adams et al., 2014).

5 Immunohistochemical analysis of TLE1 expression in IBD patients of known *NOD2* status

5.1 Introduction

5.1.1 Immunohistochemistry

Immunohistochemistry is a powerful technique used to analyse protein expression and localisation in tissues. It has previously been used to analyse the role of proteins implicated in IBD including TNFα, IL17A and SIRTUIN (SIRT1) (Honzawa et al., 2014; Melhem et al., 2015; Murch, Braegger, Walker-Smith, & MacDonald, 1993). SIRT1 is an ER stress regulator that has been implicated in the development of colitis (Melhem et al., 2015). Colonic biopsies showed decreased SIRT1 expression in the colonic crypts of UC patients when compared to non IBD controls (Melhem et al., 2015).

The IHC protocol involves tissue collection, fixing, embedding in paraffin (for storage), deparaffinisation (for analysis), antigen retrieval, staining with an antibody to the protein of interest and analysis of this staining using microscopy. Antibody staining consists of two stages, staining with a primary antibody (in this case anti TLE1) and staining with a secondary antibody linked to horseradish peroxidase (HRP). The HRP enzyme mediates a chemical reaction using diaminobenzene (DAB) as a substrate. The addition of DAB in the final stages of the IHC experiment results in brown staining in regions of tissue where HRP is present.

5.1.2 Immunohistochemical analysis of NOD2 expression

As discussed in the chapter 1, NOD2 is a pattern recognition receptor and a known IBD susceptibility gene. Mutations in *NOD2* have been associated with CD and a yeast-2 hybrid assay showed NOD2 interacts with TLE1. Initial evidence indicated that the *TLE1* rs6559629 risk increases the effect of these mutations in CD, however subsequent work has not replicated this finding (Hugot et al., 2001; Nimmo et al., 2011; Y Ogura et al., 2003).

In the intestine, NOD2 expression was first detected in Paneth cells and was originally thought to be exclusively expressed in these cells (Y Ogura et al., 2003). The absence of a functioning NOD2 antibody has meant that a majority of work has looked at mRNA expression levels of *NOD2*, however recent advances suggest that NOD2 is also expressed at the protein level in intestinal epithelial cells and monocyte derived cells (dendritic cells) of the immune system (Barnich, Aguirre, et al., 2005; Hu & Peter, 2013a).

Pattern recognition receptors such as NOD2 are generally thought to be cytoplasmic. However, recent evidence indicates that NOD2 can also shuttle to the nucleus, where TLE1 is primarily expressed (Zurek et al., 2012).

5.1.3 Immunohistochemical analysis of TLE1 expression

TLE1 expression has been well studied in a number of different cancers. Synovial sarcomas section show replicable, strong nuclear staining and TLE1 has been proposed as a diagnostic marker to differentiate synovial sarcoma from other tumors (Knösel et al., 2010). TLE1 has also been shown to stain other tumor tissues to a much lesser extent, for example malignant peripheral nerve sheath tumors (MPNT's) and Ewing's sarcoma (Kosemehmetoglu et al., 2009). Although TLE1 expression has not been analysed in normal intestinal tissue or IBD tissue, it has been shown to be expressed in gastrointestinal stromal tumors (Kosemehmetoglu et al., 2009).

In healthy tissues TLE1 is expressed in a range of cell types including endothelial cells, keratinocytes and adipocytes (Kosemehmetoglu et al., 2009). The Human Protein Atlas suggests that TLE1 is expressed in most healthy tissues, including the colon where strong nuclear expression can be observed (Uhlen et al., 2015).

In addition to nuclear staining TLE1 has also been shown the be expressed in the cytoplasm in HEK293 and HeLa cells, where NOD2 is primarily seen (Nimmo et al., 2011; Stifani, Blaumueller, Redhead, 1992).

5.2 Aims

As previously discussed, TLE1 interacts with the known IBD susceptibility gene NOD2. The general aim of this chapter was to analyse the expression of TLE1 in intestinal biopsies from healthy controls and IBD patients with and without CD associated *NOD2* variants. Specific focus was placed on cells of the intestinal crypts as Paneth cells located in the intestinal crypts show highest levels of NOD2 expression. More precisely, the aims of this work were to:

1. Identify and characterise an anti-TLE1 antibody
    a. Optimise IHC staining protocols for TLE1
    b. Replicate TLE1 staining seen in the literature
2. Determine TLE1 expression patterns in healthy ileal tissue.
    a. Analyse the expression and cellular localisation of TLE1 in the intestinal crypts of healthy ileal tissue
3. Analyse TLE1 expression in CD patients and healthy controls
    a. Quantify TLE1 expression in the ileal crypts of healthy controls and CD cases, identify any differences in TLE1 expression and localisation.
4. Analyse TLE1 expression in CD patients with and without CD associated *NOD2* variants
    a. Determine whether there are any differences in TLE1 expression or localisation in the intestinal crypt cells of CD cases with and without *NOD2* variants.

5.3 Results

5.3.1 Identification and optimisation of an anti-TLE1 antibody

Two anti-TLE1 antibodies were tested for use in this work, M101 and N18 (Santa Cruz). Ileal pinch biopsies were used for optimisation experiments, these had been collected as part of previous work conducted by Dr Craig Stevens and Dr Paul Henderson. Slides were cut from paraffin embedded sections by the pathology department at the Western General Hospital.

A standard IHC protocol was used for both optimisation and further experiments. Slides were rehydrated in glass tanks containing xylene (5 mins, twice), 100% ethanol (5 mins, twice) and 70% ethanol (5 mins, twice). Following these washes the slides were placed in a tank of distilled water. Cells were then subjected to antigen retrieval using either citric acid or trypsin. For citric acid antigen retrieval, slides were placed in a 1.5L beaker containing 10mM citric acid solution, pH6.8 (citric acid monohydrate, Sigma). Slides were heated in a 200W microwave for 20 minutes and then cooled on the bench. For trypsin treatment, slides were placed in a glass slide holder in 0.5% trypsin solution at 37˚C for 20 minutes and cooled on the bench. Once slides had cooled they were washed in a glass tank filled with water. Hydrogen peroxide solution (3%) was prepared in distilled water, slides were incubated in 300mls solution at room temperature on a rocking platform. Slides were washed in PBS and then assembled in a Sequenza slide rack (Thermo-Fisher). Slides were incubated in rabbit serum (1:5, 10 mins) (Sigma). Slides were then incubated in 100µl of antibody/rabbit serum/PBS (negative control) overnight at 4˚C and the following day washed three times in PBS. Secondary antibody was applied using a PoLink-2 plus HRP detection kit (GBI labs) using standard protocols. Following incubation with DAB, slides were washed in running water. Harris haemotoxylin (Sigma) was applied as a counterstain for 30 seconds, following this slides were washed in: running water, 68 mM lithium carbonate solution (30 secs) and running water. Finally, samples were dehydrated in glass tanks containing: 70% ethanol (twice), 100% ethanol (twice) and xylene (twice). Slides were mounted individually using Pertex mounting medium (HistoLab), air dried and analysed using a Zeiss Axioplan 2 microscope at the specified magnifications.

Optimisation of Antigen retrieval

The M101 and N18 anti-TLE1 antibodies were used at concentrations of 1:10, 1:20 and 1:50 with citric acid antigen retrieval. These concentrations were determined by analysis of the relevant literature. No staining was seen in negative control slides (secondary antibody only) or in any slides stained with the N18 antibody, strong nuclear staining was achieved with the M101 antibody at a 1:10 dilution. (Figure 42).

Trypsin antigen retrieval was used at a wider concentration of 1:20-1:500 as it had not been used in previous TLE1 IHC studies. No staining was seen in negative controls (secondary antibody only) or in sections stained with the N18 antibody. Staining was observed at the 1:20, 1:50 and 1:100 M101 anti-TLE1 antibody concentrations (Figure 43). Strong cytoplasmic and nuclear staining was observed at all three concentrations, this is in contrast to strong nuclear and weak cytoplasmic staining described in other tissues the literature. As the observed staining was determined to have high background and non- specific staining, citric antigen retrieval was used in further experiments. This conclusion was confirmed by the pathologist, Professor Donald Salter.

As staining was not achieved using two antigen retrieval methods the N18 antibody was not used in any further work. The M101 antibody was used in these experiments as optimisation experiments showed strong nuclear reactivity, as expected. Furthermore, the M101 antibody has been used extensively for IHC in previous published work involving analysis of TLE1 expression in synovial sarcoma and Ewing's sarcoma (Foo, Cruise, Wick, & Hornick, 2011; Jagdis, Rubin, Tubbs, Pacheco, & Nielsen, 2009; Knösel et al., 2010; Lino-Silva, Flores-Gutiérrez, Vilches-Cisneros, & Domínguez-Malagón, 2011; Rekhi, Basak, Desai, & Jambhekar, 2012).

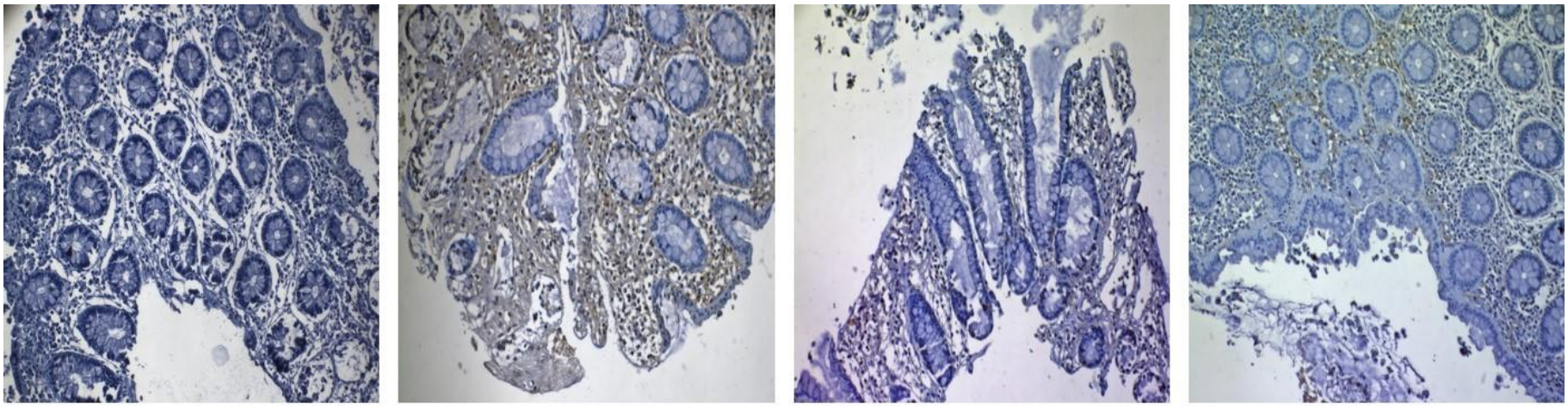


Figure 42 Optimisation of M101 anti- TLE1 antibody using citric acid antigen retrieval
The M101 anti-TLE1 antibody was used at concentrations of 1/10, 1/100 and 1/50 following citric acid antigen retrieval. A negative control image is shown on the right (secondary antibody only). Sections are DAB stained, counterstain is haematoxylin. Images have been taken at 20X magnification.

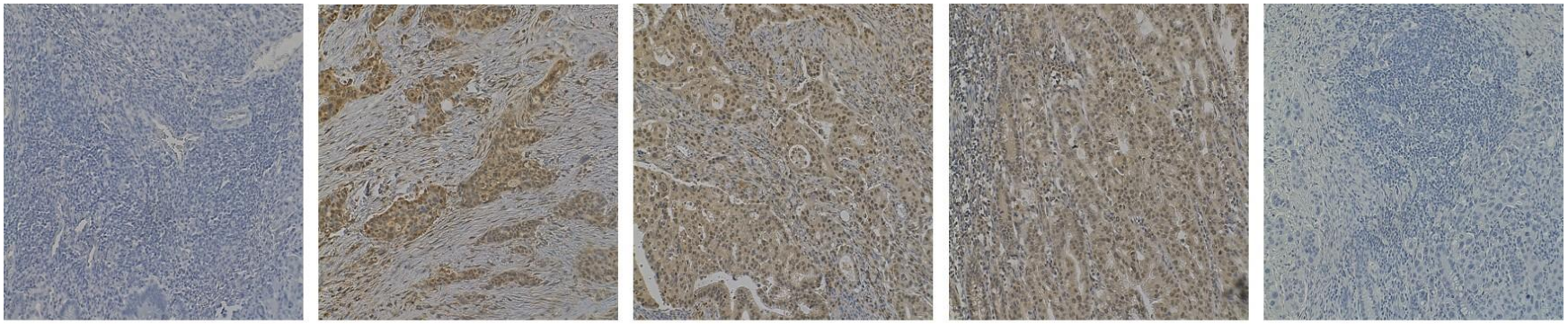


Figure 43 Optimisation of M101 anti-TLE1 antibody using trypsin antigen retrieval
The M101 antibody was used on ileal pinch biopsy slides at concentrations of 1/20, 1/50 and 1/100 following trypsin antigen retrieval. A negative control image is shown in the right (secondary antibody only). Sections are DAB stained, counterstain is haematoxylin. Images are taken at 20X magnification.

5.3.2 Replicating TLE1 staining observed in published work

A literature search was used to identify published work which had analysed TLE1 expression by IHC. In order to further characterise the M101 anti-TLE1 antibody, the tissues identified were analysed for TLE1 staining.

As discussed, synovial sarcoma tissue has been analysed for TLE1 staining in many different research papers. Five synovial sarcoma slides from different patients were obtained from the pathology service at the Western General Hospital with the kind help of Dr Catherine Black and Dr Andrew Wood. Each slide was stained according to standard protocols described above. Citric acid antigen retrieval was used and the M101 anti-TLE1 antibody was used at a concentration of 1:10. Figure 44A shows negative control slides (secondary antibody only) do not show any staining for TLE1 at 20X, 40X or 60X magnification. Figure 44B shows a synovial sarcoma slide stained with M101 anti-TLE1 antibody (1:10) positive nuclear staining can be seen in addition to very weak cytoplasmic staining. Figure 44C shows synovial sarcoma slides incubated with rabbit serum (1:20) overnight in order to determine whether the secondary antibody bound to any non-specific proteins in rabbit serum. Weak, diffuse primarily cytoplasmic staining can be seen. These staining patterns see using the M101 anti-TLE1 antibody replicate those seen in the literature.

Ewing's sarcoma tissue was used to further characterise the M101 anti-TLE1 antibody. Research shows that although some staining is observed in Ewing's sarcoma the degree of staining is much lower than that seen in synovial sarcoma. Slides from two Ewing's sarcoma cases were obtained from resources at the Western General Hospital. Figure 45A shows Ewing's sarcoma tissue negative controls, stained with secondary antibody alone, no staining is observed at either 20X, 40X or 60X magnification. The brown staining observed was present on the slides prior to deparrafinisation and was a result of sample processing by pathology, it does not represent positive staining. Ewing's sarcoma sections stained with M101 anti-TLE1 antibody are shown in Figure 45B, some nuclear staining is observed, however this is to a lesser extent to that seen in synovial sarcoma. Cytoplasmic staining is also visible in Ewing's sarcoma slides

stained with M101 anti-TLE1 antibody. At 60X magnification very diffuse staining cytoplasmic staining can be seen, no nuclear staining is visible.

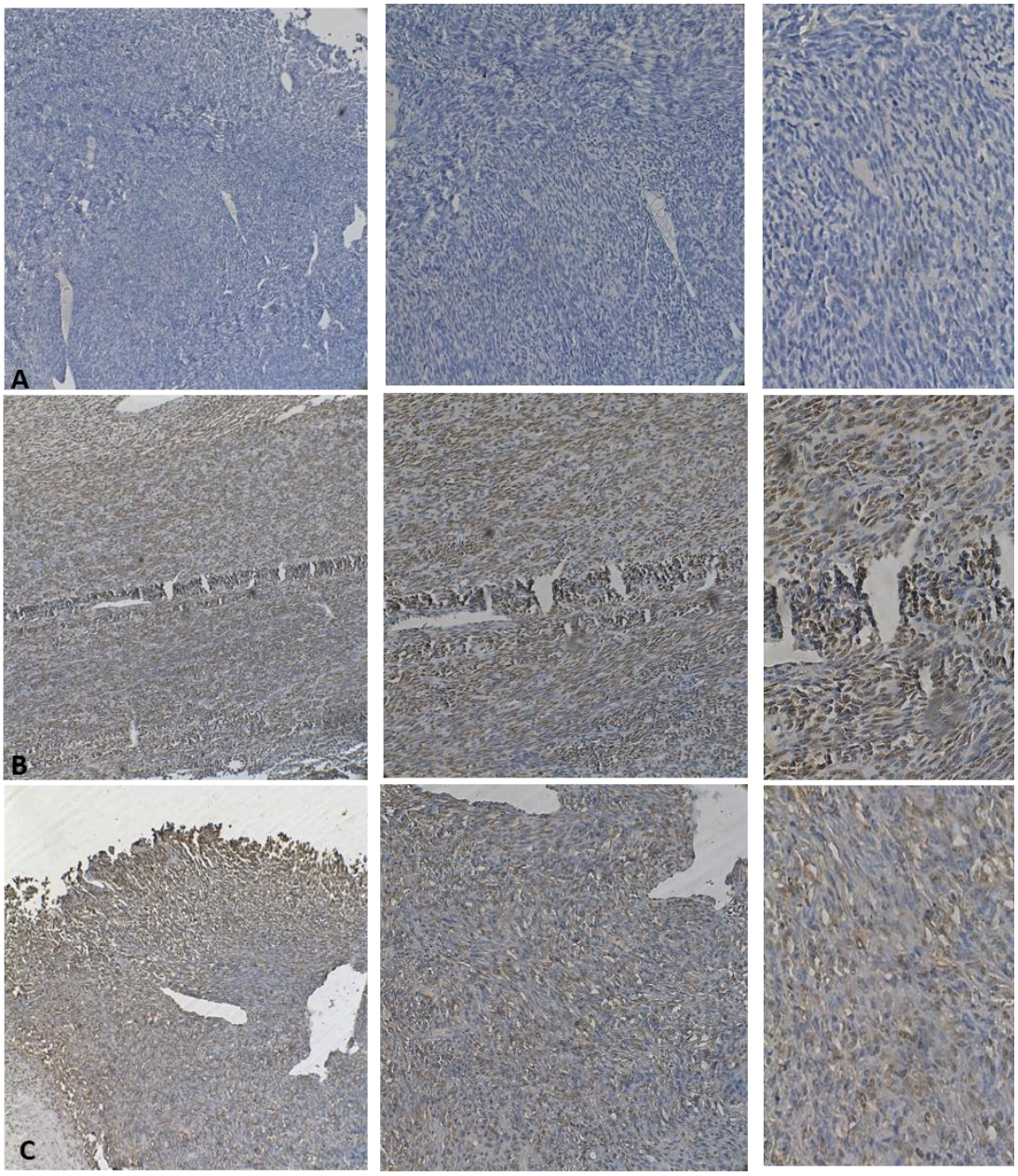


Figure 44 TLE1 expression in synovial sarcoma
A) Synovial sarcoma tissue, negative control (secondary antibody only) at 20X, 40X and 60X magnification (left to right). B) Synovial sarcoma tissue stained with M101 anti-TLE1 antibody (1/10) 20X, 40X and 60X magnification (left to right). C) Synovial sarcoma tissue incubated with rabbit serum (1/10) 20X, 40X and 60X magnification (left to right). Sections are DAB stained, counterstain is haematoxylin.

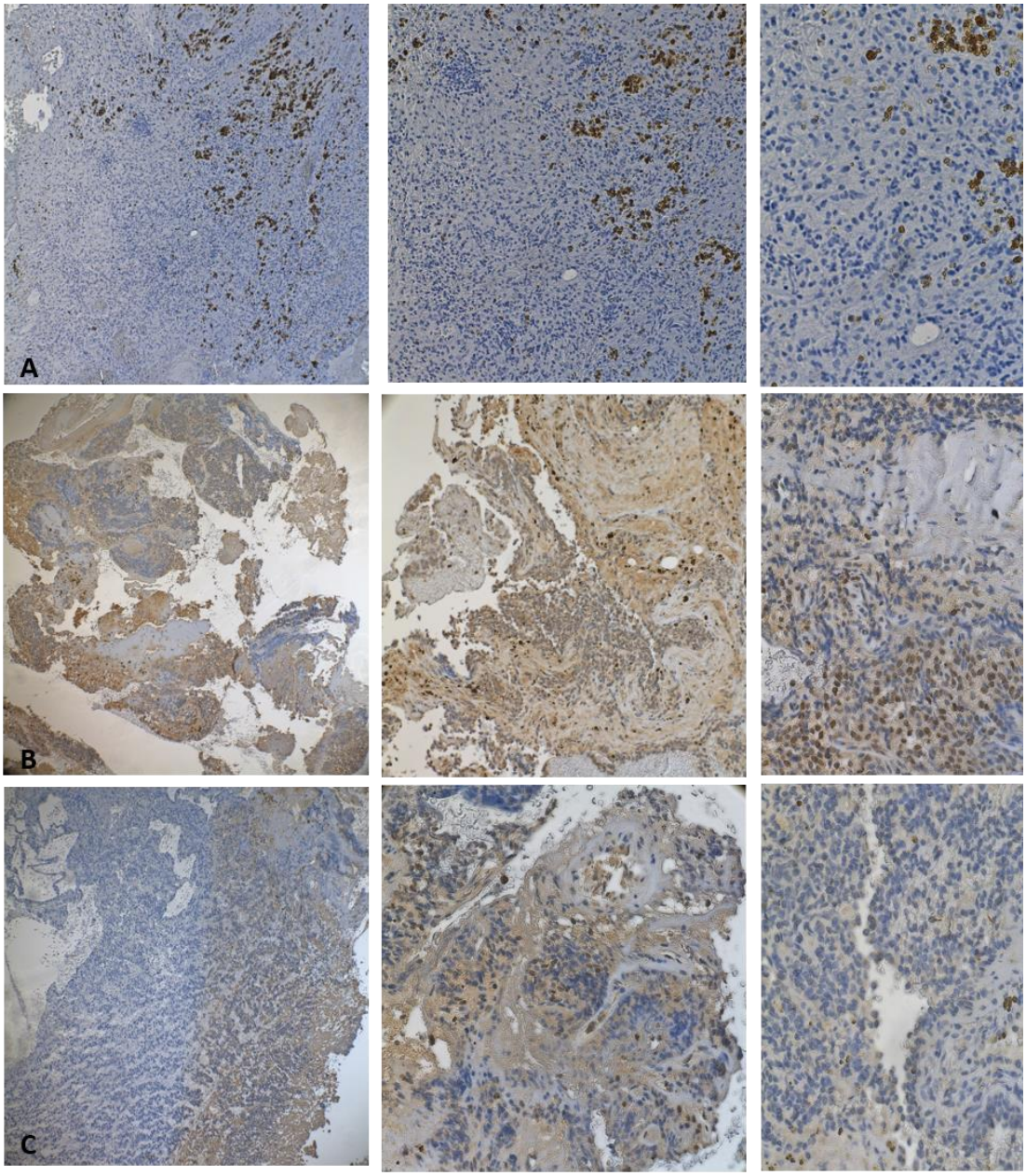


Figure 45 *TLE1* expression in Ewing's sarcoma
A) Ewing's sarcoma tissue negative control (secondary antibody only) at 20X, 40X and 60X magnification (left to right). B) Ewing's sarcoma tissue stained with M101 anti-TLE1 antibody (1/10) 20X, 40X and 60X magnification (left to right). C) Ewing's sarcoma tissue incubated with rabbit serum (1/10) 20X, 40X and 60X magnification (left to right). Sections are DAB stained, counterstain is haematoxylin.

5.3.3 TLE1 staining in healthy human tissue

TLE1 expression was analysed using a paraffin tissue panel purchased from Biochain. The array was analysed using standard IHC protocols as described above. Staining was quantified as low (0-30% of positive cells), moderate (30-60% positive cells and highly positive (60%+ positive cells). Cytoplasmic and nuclear staining were quantified separately. Tissues analysed using this panel were: heart, brain, liver, kidney, lung, pancreas, spleen and skeletal muscle (Figure 47). The age and sex of these samples is shown in

Table 14. Varied staining was seen across different tissues, pancreatic tissue showed the highest levels of nuclear expression, all other tissues showed low nuclear expression levels. Brain tissue had the highest cytoplasmic expression, expression in other tissues as low or moderate with heart, lung, spleen and skeletal tissue showing the lowest expression

Table 14).

Tissue	Sex	Age	Nuclear staining	Cytoplasmic staining
Heart	F	61	Low	Low
Brain	F	87	Low	High
Kidney	F	66	Low	Moderate
Liver	M	71	Low	Moderate
Lung	M	50	Low	Low
Pancreas	M	71	Moderate	Moderate
Spleen	M	24	Low	Low
Skeletal muscle	M	24	Low	Low

Table 14 TLE1 staining in healthy tissues

Healthy tissues were stained for TLE1 protein expression. Sex, age, clinical diagnosis and extent of staining in the nucleus and cytoplasm are shown. Low staining is classified as less than 30% positive, moderate is between 30 and 60% and high is over 60%.

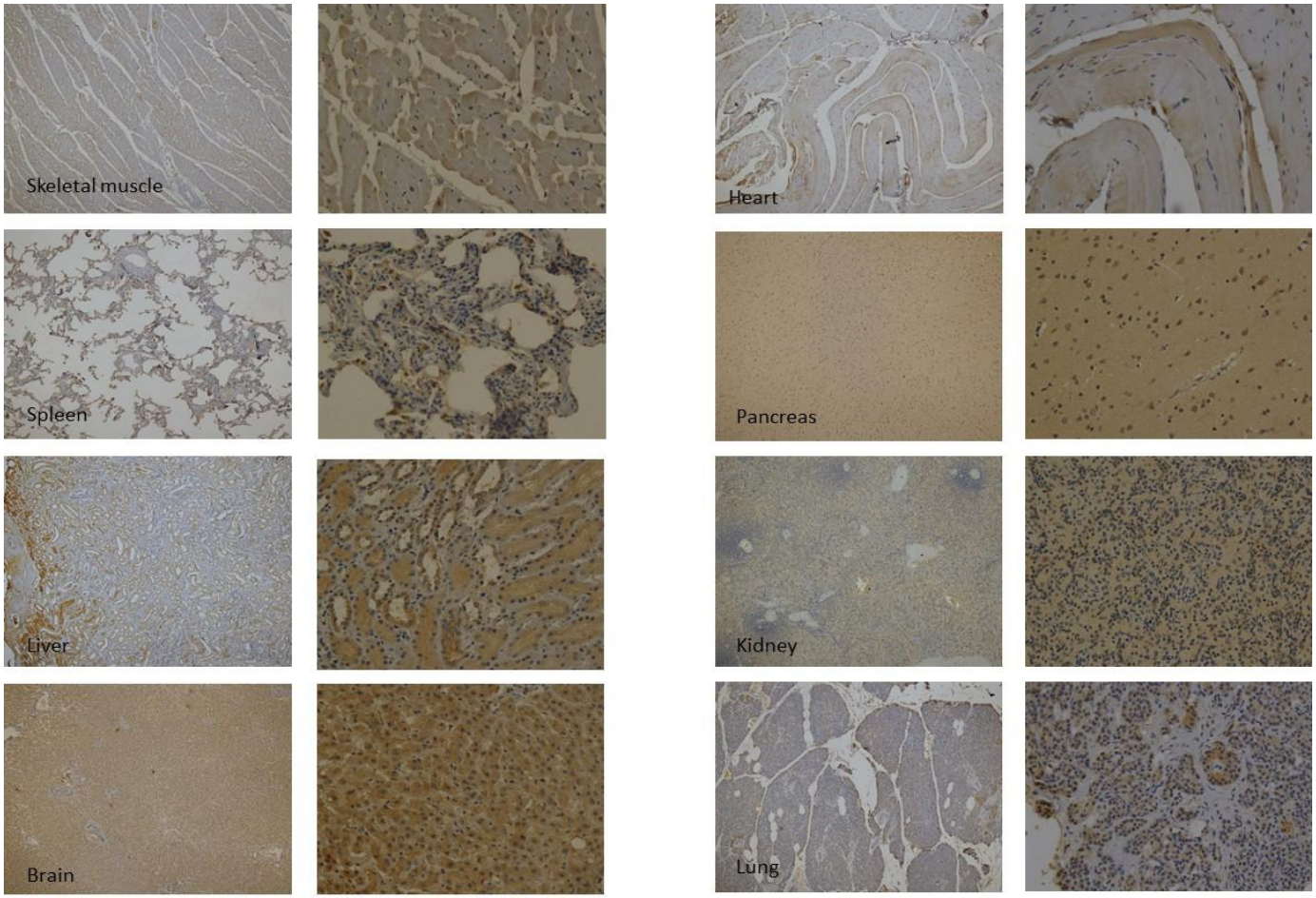


Figure 46 TLE1 expression in healthy tissues

TLE1 expression in human skeletal muscle, spleen, liver. Brain, heart, pancreas, kidney and lung. Images are shown at 10X and 40X magnification. Sections are DAB stained, counterstain is haematoxylin.

5.3.4 TLE1 staining in ileal resections from healthy controls and CD patients of known NOD2 status

Ileal resection samples were requested from the pathology department of the Western General Hospital for two healthy controls and twenty IBD patients of known *NOD2* status (10 with *NOD2* variants, 10 without *NOD2* variants) with the help of Dr Catherine Black and Dr Andrew Wood. Paraffin embedded samples were processed according to standard hospital protocols and cut into 4µM slides by the pathology department.

Prior to starting the experiment, slides from all groups (healthy controls, CD with no *NOD2* variants, CD with *NOD2* variants) were randomised and numbered with the help of another scientist. This was to blind the experiment and prevent any unconscious bias during both the experiment and the analysis. Each slide was subject to TLE1 staining as described above and analysed using a Zeiss Axioplan 2 microscope. Five frames from each slide were imaged at both 20 and 40 times magnifications and grouped according to TLE1 expression (low 0-30% of positive cells, moderate 30-60% positive and highly positive 60%+). TLE1 expression was quantified in 100 cells, using images taken at 40X magnification. Cytoplasmic and nuclear staining were quantified separately. All images analysed consisted of approximately 50% intestinal crypts as this region was of particular interest in this study. No slides showed overall staining higher than 85%. This protocol was has been used in other immunohistochemistry/ IBD studies and was advised by Professor Donald Salter.

5.3.5 TLE1 is highly expressed in the ileal crypts of healthy controls

TLE1 expression was analysed in ileal resections from two healthy controls. The exact details of diagnosis are not available to us under the conditions of the relevant ethical review, however they were confirmed to be uninflamed, non-cancerous and non IBD samples. Representative images of TLE1 staining in these samples is shown in Figure 47. Strong nuclear and limited cytoplasmic staining was observed in both healthy controls (A= 90%, B-85% of cells in the intestinal crypts). This staining was observed in the nucleus and cytoplasm of a majority of Paneth cells analysed i.e. 74% of Paneth cells analysed were positive for TLE1 staining.

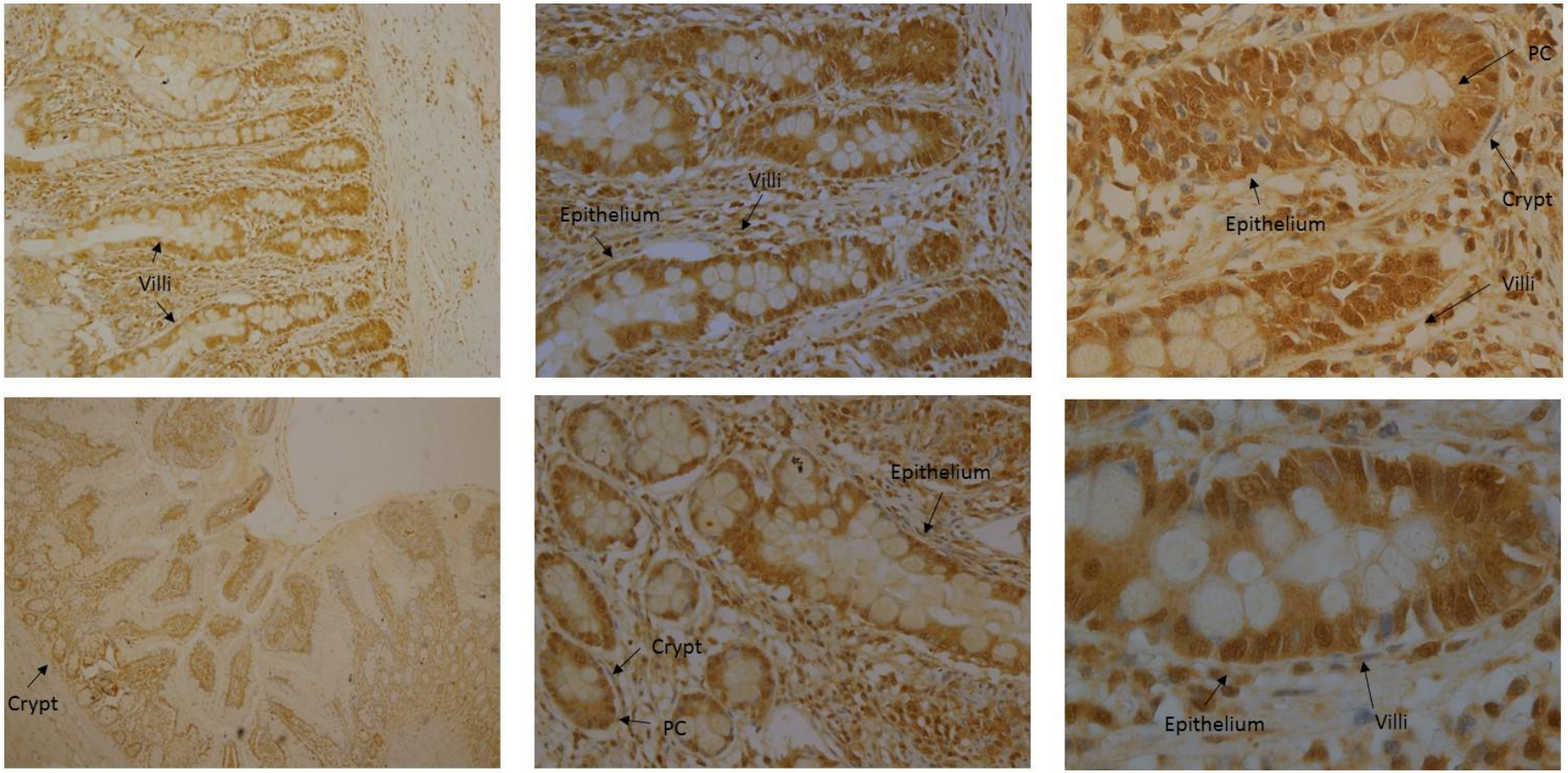


Figure 47 TLE1 expression in the ileal tissue of healthy controls

TLE1 staining in healthy ileal tissue obtained from two healthy controls. The top panel represents images taken from one healthy control, from left to right, 10X, 20X and 40X images showing villi and crypts respectively. The lower panel shows the same for the second healthy control. Sections are DAB stained, counterstain is haematoxylin.

5.3.6 TLE1 shows varied expression in CD patients with and without CD associated NOD2 variants

TLE1 expression was analysed as described above in ileal resections from 10 patients with CD associated *NOD2* variants and 10 patients without these *NOD2* variants. Overall, there was varied staining ranging from completely negative to highly positive (Figure 48A, Figure 49A and Figure 50A). TLE1 expression was observed in both intestinal epithelial cells and Paneth cells of the intestinal crypts. Paneth cells in samples with moderate or high TLE1 expression showed nuclear and cytoplasmic staining (Figure 49 A and Figure 50A). Nuclear staining was also seen in the intestinal epithelial cells of the intestinal crypt (Figure 49) and in the lamina propria (Figure 48B).

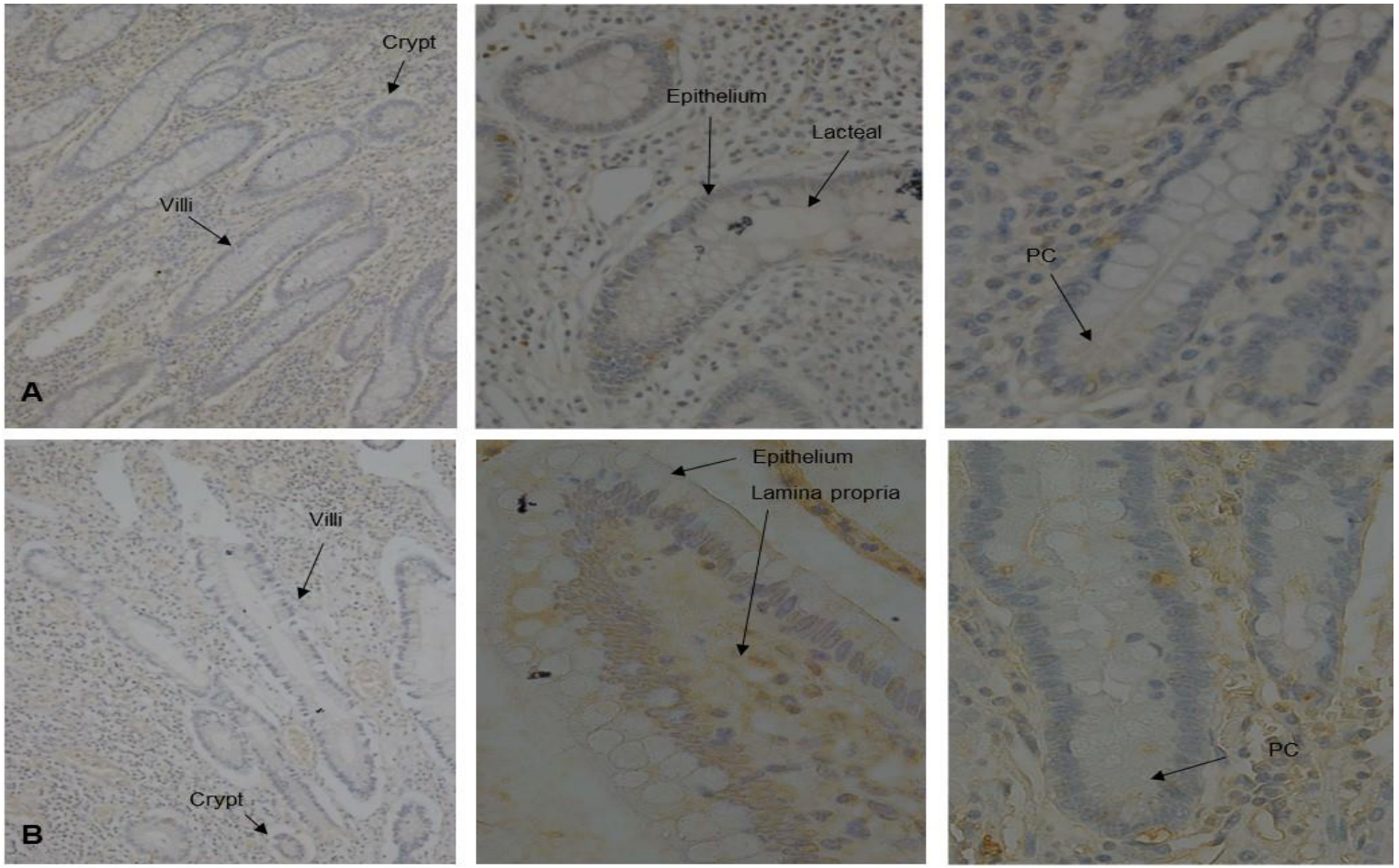


Figure 48 A subset of ileal CD patients with and without CD associated *NOD2* variants show low levels for TLE1 staining

Four CD patients with mutant NOD2 and three CD patients with WT NOD2 showed little to no TLE1 staining in any areas of the ileum. Sections are DAB stained, counterstain is haematoxylin. From left to right, image taken at 10 times magnification, image of villi at 40 times magnification and crypt taken at 40 times magnification.

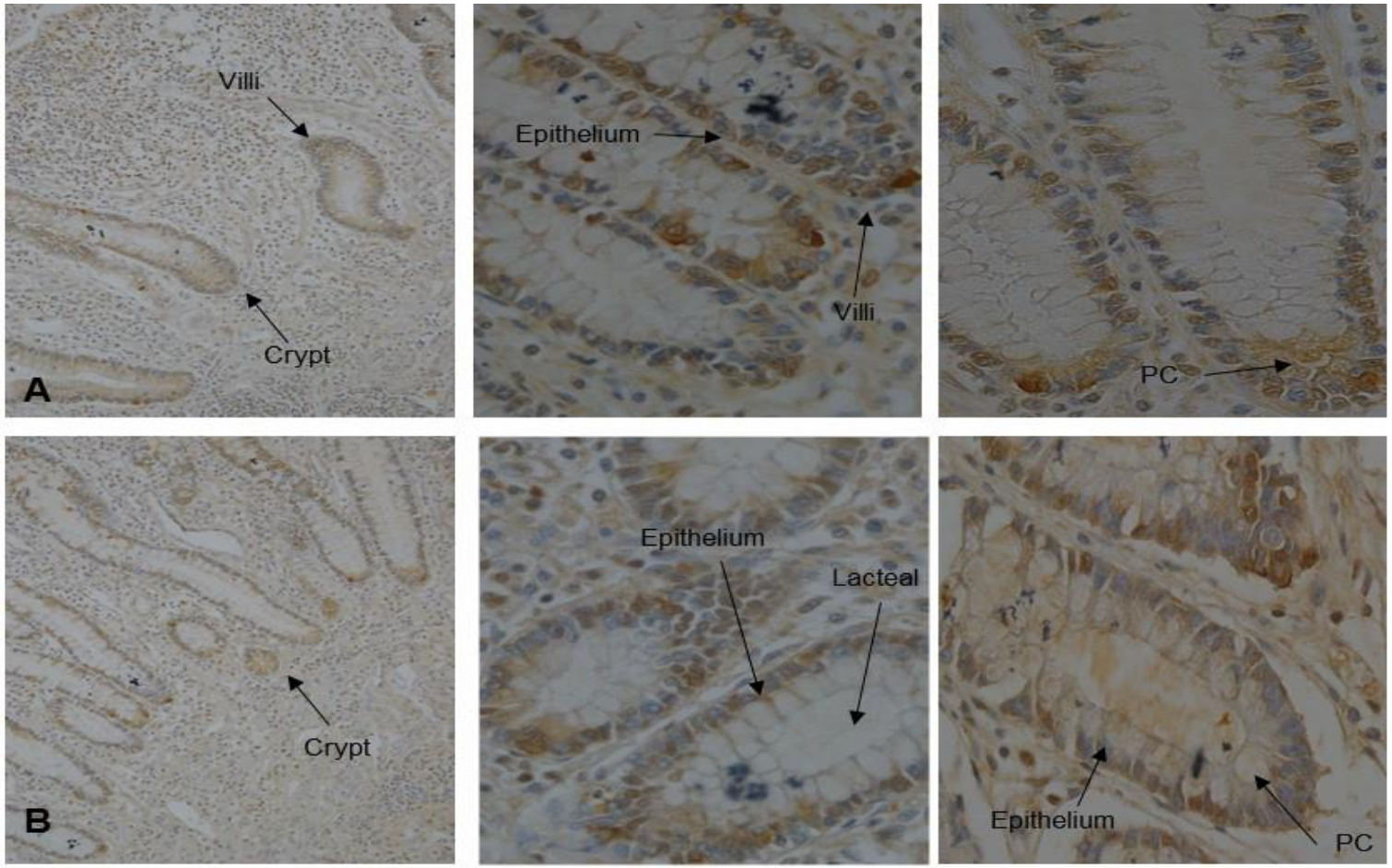


Figure 49 A subset of Ileal CD patients with and without CD associated *NOD2* variants exhibit moderate TLE1 staining

Four CD patients with mutant NOD2 and six CD patients with WT NOD2 showed between 20-50% of ileal cells staining positively for TLE1. Sections are DAB stained, counterstain is haematoxylin. From left to right, image taken at 10 times magnification, image of villi at 40 times magnification and crypt taken at 40 times magnification.

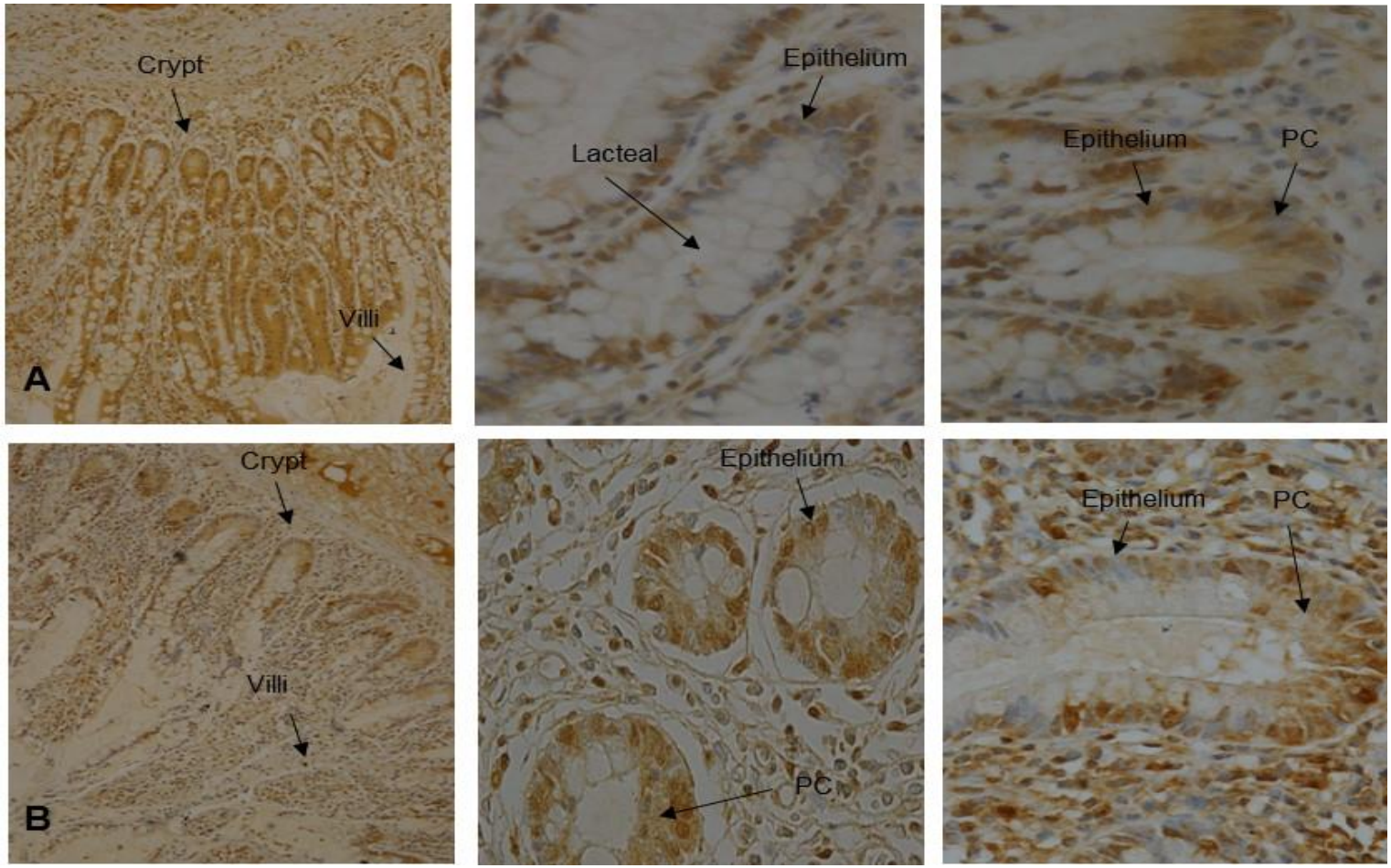


Figure 50 A subset of Ileal CD patients with and without CD associated *NOD2* variants exhibit highly positive TLE1 staining

Four CD patients with mutant NOD2 and six CD patients with WT NOD2 showed over 50% of ileal cells staining positively for TLE1. Sections are DAB stained, counterstain is haematoxylin. From left to right, image taken at 10 times magnification, image of villi at 40 times magnification and crypt taken at 40 times magnification

5.4 Discussion

In this work, I have optimised an anti-TLE1 antibody and used it to analyse TLE1 expression in the ileum of healthy controls and CD cases. The CD patients used in this study were of known *NOD2* status in order for us to further investigate the TLE1/NOD2 interaction and whether NOD2 mutations affect TLE1 expression or localisation. This is the first study to show the expression patterns of TLE1 in human ileal tissue.

Overall healthy controls showed strong TLE1 staining in the intestinal crypts, expression was seen in both intestinal epithelial cells and Paneth cells. Both the NOD2 mutant and NOD2 wild type ileal resections showed varied TLE1 staining; ranging from less than 20% positive to 85% positivity. Sections with moderate or highly positive TLE1 staining showed expression of TLE1 in Paneth and intestinal epithelial cells.

The cause of the variation seen in TLE1 expression in CD cases regardless of genotype is not known. The sections analysed were all from uninflamed cases so this is unlikely to be the cause. Additional information on medical treatment and IBD markers for these patients e.g. fecal calprotectin would of interest to future work in order to determine whether TLE1 expression correlates with active disease, disease relapse or specific treatments. The relevant ethical review did not allow for information on the type of *NOD2* mutation. As the L1007fsInsC mutation does not localize to the plasma membrane it would be interesting to analyse each of the mutations individually with respect to TLE1 expression and localisation in future work.

Recent work in our laboratory, conducted by Helen Newbery, had identified a functioning NOD2 antibody. NOD2 expression is primarily in the Paneth cells of healthy control resections and CD patients (regardless of *NDO2* status), this finding is line with the literature (Y Ogura et al., 2003). There are isolated epithelial cells that

appear to show cytoplasmic positivity. Overall NOD2 protein expression in the intestinal crypts is much lower than TLE1 expression. Although expression of TLE1 and NOD2 in the same cell types makes the interaction observed in cell lines more plausible it does not provide definitive evidence that NOD2 and TLE1 interact in the human ileum. Analysis of co-localisation of TLE1 and NOD2 was not possible in this study as both TLE1 and NOD2 are raised in rabbit. Cross reactivity between antibodies of the same species means that any results from these experiments would be uninterpretable. Current work in our laboratory is focused upon analysing expression of proteins in serial sections taken from the same sample. This may allow for analysis of TLE1 and NOD2 expression in the same cells. Future work may could focus on identification of alternative TLE1 antibodies raised in different species allowing analysis of co localisation of NOD2 and TLE1 with the use of fluorescently tagged secondary antibodies. Alternatively, there are IHC techniques that can be used when antibodies are raised in the same species. Martinez et al used antibodies raised in the same species to analyse the protein adrenomellulin and glucagon in the rat pancreas. The protocol used in this work involves a stripping technique whereby slides are stained with one antibody and staining is visualised, the slide is then stripped for one minute in 2.5% $KMnO_4$/ 5% H_2SO_4/ distilled H_2O, 1:1:30 (vol/vol) and 0.5% Na metabisulfite in distilled water for 3 min. Following stripping slides are washed in water for two hours prior to re staining with the second primary antibody. Controls showed no signal from stripped sections stained with secondary antibody alone which suggests that the stripping was effective (Martínez, Cuttitta, & Teitelman, 1998). This technique may be worth trying if an alternative functioning TLE1 antibody cannot be found.

As discussed, the differences in TLE1 expression between CD patients could not be attributed to *NOD2* genotype, even once the specific type of mutation is known. It may be the case that other proteins are involved in the NOD2/TLE1 interaction i.e. they form a complex. Both NOD2 and TLE1 appear to be hub proteins; they are involved in many different signaling pathways and interact with hundreds of other proteins. It may also be of interest to use isolate the TLE1/NOD2 potential complex and use a combination of mass spectrometry and peptide sequencing to identify the proteins involved.

6 Sequencing *TLE1* in Crohn's disease patients and healthy controls.

6.1 Introduction

SNPs are single base pair changes which can be used as markers in Genome Wide Association (GWA) studies. They tag blocks of linkage disequilibrium, these are regions of DNA that are not separated by recombination. Association of a SNP with a particular phenotype or disease has been used to identify regions of DNA that may contribute to pathogenesis. The majority of the 163 IBD susceptibility loci identified have been found as a result of GWA studies, to date, 71 susceptibility loci for CD have been discovered by GWA studies (Franke et al., 2010; Jostins et al., 2012).

In 2011, Nimmo et al typed eight SNPs in the TLE1 gene and showed that three of these SNPs were significantly associated with CD in a combined Scottish cohort. The eight SNPs were genotyped in two independent Scottish cohorts from Edinburgh (controls n= 841, CD n=352) and Dundee (controls n=674, CD n=296). Three SNPs were significantly associated with CD in the Edinburgh cohort: rs11139315 (p=0.01), rs2796469 (p=0.04), rs6559629 (p=0.02). In the Dundee cohort: rs2796469 (p=0.04), rs6559629 (p=0.09), rs10867783 (p=0.0005), rs7856583 (p=0.02) were significantly associated with Crohn's disease. When these two cohorts were combined three SNPs remained significantly associated with Crohn's disease: rs11139315 (p=0.003), rs2796469 (p=0.004), rs6559629 ($p= 4\times10^{-4}$), MAF's and p values for all *TLE1* SNPs are shown in . The location of the eight *TLE1* SNPs typed in this study is shown in Figure 52, the three SNPs that showed statistically significant differences in allele frequency are highlighted. Two of these SNPs: rs2796469 and rs6559629 tag the same LD block, as shown in Figure 51.

Analysis of the *NOD2* and *TLE1* genotype frequencies in the aforementioned cohort showed that the rs6559629 *TLE1* risk allele appears to increase the effects of *NOD2* mutations. In patients with only a *NOD2* risk allele there was an odds ratio of 2.49, confidence interval 0.61-10.14, whereas in the patients with one *TLE1* risk allele and a *NOD2* risk allele an odds ratios of 10.79, confidence interval 3.01-38.64 were observed (Table 16). These results have not been replicated in data from larger, UK IBD cohorts, suggesting this finding is only relevant to the Scottish population or it was an artefact of the smaller cohort used in this study (controls n=630, CD n=280).

	Controls n=1515	All CD (n = 648)		L1 (n = 170)		L2 (n = 188)		L3 (n = 171)	
TLE1	%MAF	%MAF	p value	%MAF	p value	%MAF	p value	%MAF	p value
Combined subphenotypic analysis of CD data by location									
rs11139315	36.5	41.3	0.003	39.7	0.24	40.4	0.14	39.7	0.24
rs2796469	37.8	33.2	0.004	27.6	2 × 10−4	36	0.49	33	0.08
rs6559629	48.5	54.5	4 × 10−4	60.6	3.1 × 10−5	48.9	0.88	56.8	0.004
rs7045812	15.2	15.7	0.64	15.1	0.96	15.2	0.99	15.4	0.92
rs10867783	43.2	39.9	0.05	40.1	0.28	36.8	0.02	43.5	0.91
rs7856583	45.5	47.5	0.23	45.6	0.99	50	0.1	50.6	0.08
rs2378591	37.6	36.8	0.6	41.1	0.21	33.3	0.11	37.4	0.92
rs3739581	24.5	25.3	0.56	28.1	0.14	22.5	0.39	26	0.39

Table 15 TLE1 SNP frequencies in combined Scottish CD cohort (adapted from (Nimmo et al, 2011))

TLE1 SNP allele frequency in healthy controls, all CD patients and patients with L1, L2 and L3 CD separately. L1, L2, and L3 describe the location of disease, refer to Table 17. Each allele as a fraction of the total number of alleles, %MAF and corresponding p values are shown. Information for the SNP with the strongest association (rs6559629) is highlighted.

	Number of risk alleles (NOD2)	Number of risk alleles (TLE1)	Controls (n=1272)	CD (n=590)	OR	CI	CD (L1) (n=162)	OR	CI
	0	0	264	106	1		21	1	0.54-1.60
	0	0	593	208	0.87	0.66-1.15	44	0.93	1.41-4.21
	0	2	232	142	1.52	1.21-2.07	45	2.44	0.44-3.35
	1	0	52	20	0.96	0.55-1.68	5	1.21	1.75-6.30
	1	1	87	56	1.6	1.07-2.40	23	3.32	2.23-10.16
	1	2	37	37	2.49	1.50-4.14	14	4.76	0.34-29.40
	2	0	4	4	2.49	0.61-10.14	1	3.14	4.68-93.79
	2	1	3	13	10.79	3.01-38.64	5	20.95	
	2	2	0	4			4		
NOD2 carriage			14.3%	22.7%			32.1%		
TLE1 carriage			74.9%	78.0%			83.3%		
NOD2 +TLE1 carriage			10.0%	18.6%			28.4%		
NOD2 MAF			7.5%	13.1%			19.1%		
TLE1 MAF			48.0%	54.5%			61.1%		

Table 16 Frequency and odds ratios of *TLE1* rs6559629 and *NOD2* risk alleles in a combined Scottish cohort

Number of *TLE1* and *NOD2* risk alleles in combined Scottish cohort, number of alleles, odds ratios and confidence intervals are shown. *NOD2* and *TLE1* cohort MAFs and percentage of the cohort with *TLE1* and *NOD2* risk alleles both individually and combined are shown.

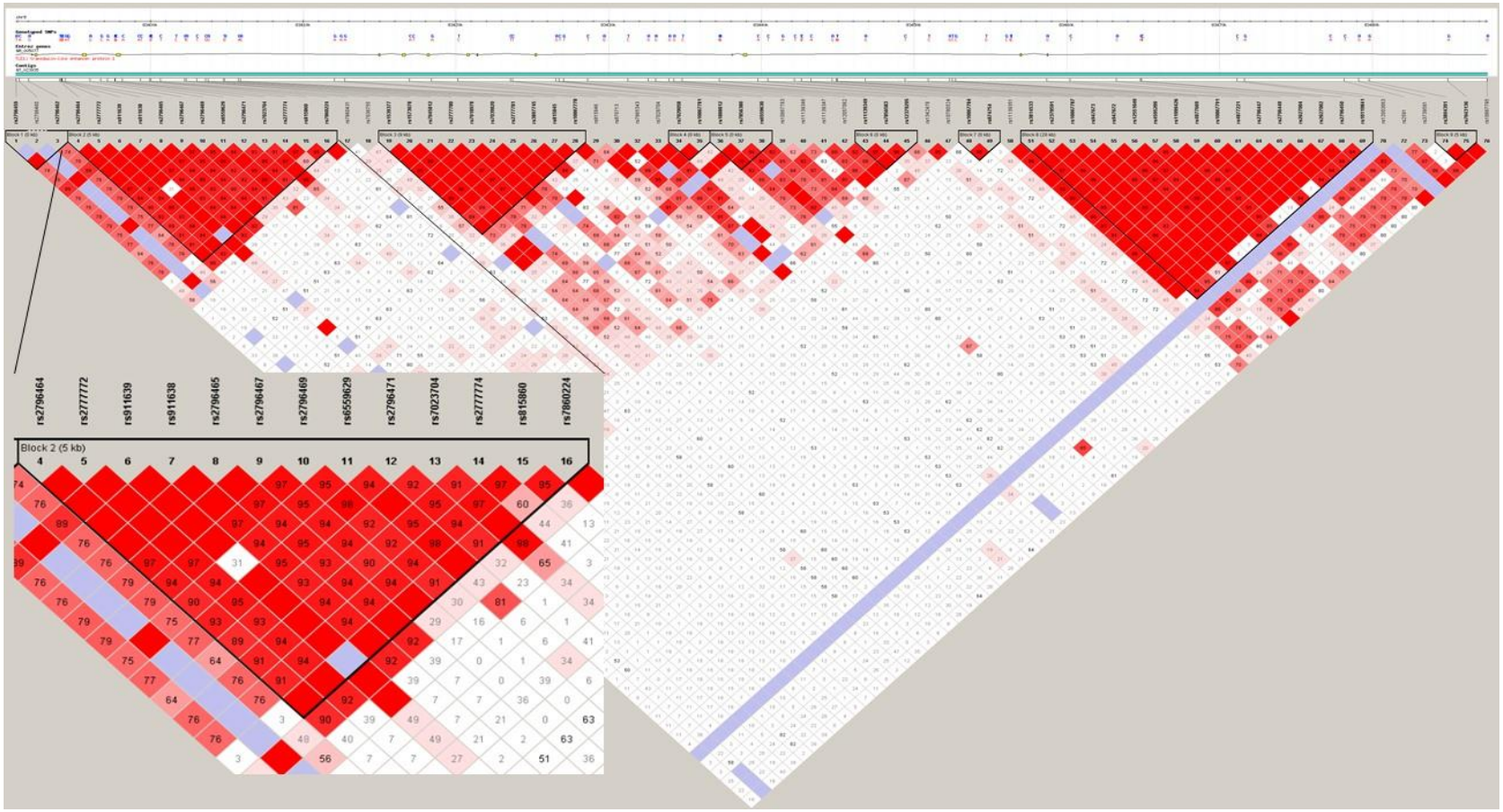


Figure 51 Linkage Disequilibrium patterns in TLE1; gene

Map of LD blocks in *TLE1* gene constructed using data from HAPMAP release#28 and Haploview v4.2. Block one contains the rs6559629 SNP, the *TLE1* SNP most strongly associated with CD ($p=4\times10^{-4}$)

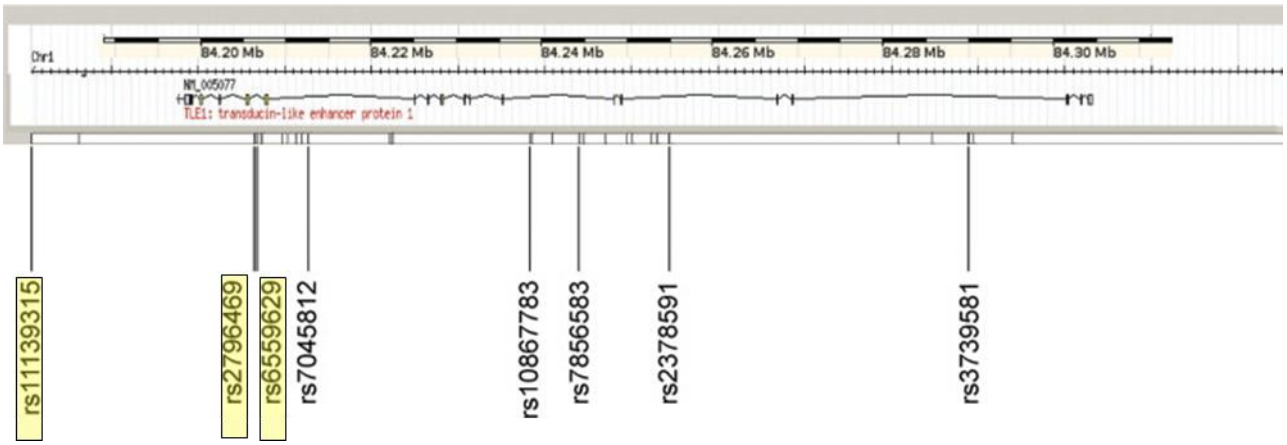


Figure 52 *TLE1* SNPs associated with CD (Nimmo et al, 2011)

The eight SNPs genotyped as part of the Nimmo et al, 2011 study are shown. The three SNPs highlighted in yellow were significantly associated with CD in a combined Scottish cohort. Diagram adapted from Nimmo et al 2011.

6.2 Aim

Initial studies showed that three SNPs in *TLE1* are associated with CD in a combined Scottish cohort (Nimmo et al., 2011). The strongest association observed was between rs6559629 and CD ($p= 4\times10^{-4}$). This aim of this chapter was to determine whether there was a causative variant underlying the association of rs6559629. In order to enrich for mutations linked to the rs6559629 association, patients with the rs6559629 risk allele and controls without the risk allele were used in this study. The aim of this chapter can be broken down into several stages:

1. Sequencing coding regions of *TLE1* in a Discovery cohort

2. Sequencing relevant non coding regions in a Discovery cohort

3. Analysing the frequency of any mutations found in a larger Scottish cohort.

6.3 Methods

6.3.1 Cohort selection

The first stage of this involved selecting a cohort of patients to sequence. A group of 24 CD patients and 24 healthy controls were selected (DNA samples from Western General Hospital stored at the Wellcome Trust Clinical Research facility, WTCRF). Average age of healthy controls was 42 and 48 for CD patients (Table 19). CD patients with ileal disease ((L1 or L3) were chosen preferentially as the rs6559629 association was shown in patients with ileal disease (Nimmo et al., 2011) .Table 17 describes the Montreal classification system and Table 18 shows the number of patients in each sub phenotypic category. These patients had previously been genotyped for the rs6559629 SNP. CD patients who were homozygous for the rs6559629 risk allele were chosen and healthy controls who were homozygous for the alternative allele were used in order to increase our chances of finding a mutation underlying the association seen between rs6559629 and CD. The cohort of 24 individuals of each phenotype allowed identifications of mutations present at a greater than 1% frequency in the population. DNA samples were provided as 20µl aliquots (2µg/ul) from samples stored at the Welcome Trust clinical research facility (WTCRF).

	Sub phenotypes
Location	L1 ileal L2 colonic L3 ileo-colonic
Behaviour	B1 non stricturing, non penetrating B2 stricturing B3 penetrating

Table 17 Montreal classification of IBD sub phenotypes and behaviours

Table describing disease location and behaviour sub phenotypes of CD, according to the Montreal classification.

	B1	B2	B3
L1	7	3	5
L2	5	0	1
L3	0	2	2

Table 18 Sub phenotypic classification of CD patients in Discovery cohort

Sub phenotypic classification of disease location and behaviour according to the Montreal classification of the 24 CD patients in the cohort used for this study.

	Average Age		Gender	
	Male	Females	Males	Females
Healthy controls	43	40	11	13
CD patients	43	53	14	10

Table 19 Age and genotype information for Discovery cohort

Average age of males and females in both males and females in the Discovery cohort. Number of male and females in controls and CD patients is also shown.

6.3.2 PCR and Sanger sequencing

Each exon and intronic region to be sequenced was PCR amplified according to standard protocols. Figure 53 shows the structure of the *TLE1* gene, each exon was amplified, purified and Sanger sequenced (ABI 3130 genetic analyser, Institute of Genetics and Molecular Medicine sequencing service, University of Edinburgh). The same primers were used for PCR and sequencing. Primers were designed using Primer 3 for 500bp regions, at either side of each exon 50bp of intronic sequence was included in the PCR region (Untergasser et al., 2012). Primers were optimised using gradient PCR (55-65℃) (Appendix 1).

6.3.3 Data Analysis

Sequencing data was analysed using Geneious v7.0.6. Statistical analysis was carried out in Excel. The Students T Test was used to generate a p value, heterozygous and homozygous

Figure 53 *TLE1* gene

Schematic diagram of *TLE1* gene, exons are shows as blue boxes separated by black lines representing each intron. The rs6559629 SNP is shown as a purple line between introns 15 and 16.

(minor allele) individuals were grouped together i.e. the presence of one copy of the minor allele was used as the "risk factor"

6.4 Results

6.4.1 Discovery Cohort

Sequencing of TLE1 exons

Primers for each of the *TLE1* exon were designed and optimised by gradient PCR (see Appendix). Additional sequencing primers used for exons greater than 450bp are shown in brackets underneath PCR primers. Prior to sequencing, 5µl of each PCR product was run on a 1.5% agarose gel to check for visible clear bands of PCR product at the correct size.

Analysis of sequencing data revealed two exonic SNPs: rs147523347, rs114633202 (**Error! Reference source not found.**). The minor allele frequencies (MAF) of each of these SNPs according to the Ensembl database (European population) were: 0.006 and 0.003, respectively. It is important to note that any p values discussed here on do not reflect differences in allele frequencies in the general population as CD patients with the rs6559629 risk allele and healthy controls without the risk allele were chosen for this study. The SNP in exon 2, was heterozygous in one healthy control and two CD patients, all other patients and controls were homozygous for the major allele (A). The rs114633202 SNP in exon 6, was heterozygous in one healthy control, all other CD patients and healthy controls were homozygous for the major allele (A) (p=0.56). The rs372238712 SNP was homozygous for the minor allele (T) in one CD patient and heterozygous in one healthy control, all other individuals were homozygous for the major allele (C) (**Error! Reference source not found.**).

Sequencing of non- coding regions

Sequencing Intron 15/16

The intron 15/16 region contains two of the three SNPs in *TLE1* that are associated with CD: rs2796469 (p=0.004), rs6559629 ($p= 4\times10^{-4}$) (Nimmo et al., 2011), hence it was sequenced to identify any mutations underlying these associations.

Sequence analysis revealed four SNPs in this region: rs7023704, rs2796471, rs6559629, rs7019824 (**Error! Reference source not found.**). All CD patients were homozygous for the minor allele (G) of the rs6559629 SNP and healthy controls were homozygous for the major

allele (A). All minor allele frequencies discussed were obtained from the Ensembl database (European population). The rs7023704 SNP had a MAF of 0.21 and was homozygous for the minor allele in on CD patient and two healthy controls, it was heterozygous in one CD patient and one CD control and homozygous for the major allele in all other individuals (p=0.65). The rs2796471 SNP has a minor allele frequency of 0.21, one CD patient and two healthy controls were heterozygous for the SNP, all other patients and controls were homozygous for the major allele (A) (p=0.65). The third allele, rs7019824 has a MAF of 0.1, one CD patient and one healthy control were heterozygous, all other individuals were homozygous for the major allele (C) (p=1).

Sequencing Intron 16/17

The majority of SNPs identified GWA studies are in non- coding regions as is the case with rs6559629, the SNP in *TLE1* most strongly associated with CD. A recent research focus has been to try and analyse the direct functional significance of mutations in non- coding regions. The SuRFr pipeline was used to identify the regions of DNA, tagged by rs6559629, that may be of particular interest to sequence (Ryan, Morris, Porteous, Taylor, & Evans, 2014). This pipeline was used with the kind help of its creator Niamh Ryan. The SuRFr pipeline uses data from the UCSC genome browser including histone acetylation and methylation, transcription factor binding sites, sequence conservation and DNase 1 hypersensitivity to predict which non coding regions are most likely to have functional significance. The LD block tagged by rs6559629, shown in block 1 of Figure 51 was used as the input DNA region. The highest scoring region in the output within this LD block was the intron 16/17 region shown in Figure 54. This region contained the largest number of SNPs of potential functional significance and also contains the start site for a predicted alternative transcript of *TLE1* in addition to histone acetylation markings and DNAse1 hypersensitivity clusters.

In total, nine variants were found in intron 16/17, one of which was a 7bp deletion (rs376377585). Six of the SNPs were present at frequencies that were statistically significantly different according to the student's t test: rs376377585 (p=0.076), rs2796466 (p=0.085), rs911638 (p=0.085), rs2777772 (p=0.0003), rs2796464 (0.0201) and rs2796465 ($p=1.93\times10^{-6}$). The MAF's of the nine SNPs ranged from 0.002-0.12

Chromosome Location (forward strand)	Exon	ID number	Alleles	MAF (minor allele)	Type of variation	CD patients Homozygous (major allele)	CD patients Homozygous (minor allele)	CD patients Heterozygous	Healthy controls Homozygous (major allele)	Healthy controls Homozygous (minor allele)	Healthy controls Heterozygous	P value
9:81687396	2	rs147523347	G/A	0.006 (G)	Synonymous Exonic SNP	23	0	2	23	0	1	1
9:81652262	6	rs114633202	A/G	0.003 (G)	Synonymous Exonic SNP	24	0	0	23	0	1	0.56

Table 20 SNPs found in TLE1 exons in a cohort of 24 CD patients and 24 healthy controls

This table details the results from sequencing intron 16/17 in the previously described cohort of 24 CD patients and 24 healthy controls. The table shows chromosomal location, SNP ID numbers, type of variation, the number of patients and controls that are homozygous for the major or minor allele, the number of patients that are heterozygous and the p value for the difference in minor allele frequency between patients and controls.
Three synonymous SNPs were found in *TLE1* exons 2, 6 and 13. Allele frequencies of these SNPs were not different between CD patients and healthy controls.

Chromosome Location (forward strand)	ID Number	Alleles	MAF (minor allele)	Type of variation	CD patients Homozygous (major allele)	CD patients Homozygous (minor allele)	CD patients Heterozygous	Healthy controls Homozygous (major allele)	Healthy controls Homozygous (minor allele)	Healthy controls Heterozygous	P value (Student's t test)
9:81592983	rs7023704	C/G	0.22 (G)	Intronic SNP	22	1	1	21	2	1	0.65
9:81592942	rs2796471	A/G	0.21 (G)	Intronic SNP	23	0	1	22	0	2	0.65
9:81592445	rs6559629	G/A	0.48 (G)	Intronic SNP	0	24	0	24	0	0	N/A
9:81592295	rs7019824	C/T	0.10 (T)	Intronic SNP	23	0	1	23	0	1	1

Table 21 SNPs found in intron 15/16 of TLE1 in a cohort of 24 CD patients and 24 healthy controls

This table details the results from sequencing intron 16/17 in the previously described cohort of 24 CD patients and 24 healthy controls. The table shows chromosomal location, SNP ID numbers, type of variation, the number of patients and controls that are homozygous for the major or minor allele, the number of patients that are heterozygous and the p value for the difference in minor allele frequency between patients and controls.
In the intron 15/16 region three SNPs in addition to rs6559629 were found: rs7023704, rs2796471, rs7019824, allele frequencies were not different between CD patients and healthy controls. The p value for rs6559629 is not applicable as CD patients with the rs6559629 risk allele and healthy controls without the risk allele were selected for the cohort.

Chromosomal Location (forward strand)	ID number	Alleles	MAF (minor allele)	Type of variation	CD patients Homozygous (major allele)	CD patients Homozygous (minor allele)	CD patients Heterozygous	Healthy controls Homozygous (major allele)	Healthy controls Homozygous (minor allele)	Healthy controls Heterozygous	P value (Student's t test)
9:81590171-81590177	rs376377585	GCACTGA/-	NA	Intronic indel	24	0	0	21	3	0	0.076
9:81590773	rs373720466	G/A	NA (A)	Intronic SNP	22	0	2	22	0	2	1.00
9:81590772	rs187303459	C/T	0.002 (T)	Intronic SNP	23	0	1	24	0	0	0.32
9:81590592	rs62578588	C/T	0.074 (C)	Intronic SNP	24	0	0	23	1	0	0.32
9:81589864	rs2796466	G/T	0.290 (T)	Intronic SNP	16	5	3	10	6	8	0.085
9:81589485	rs911638	C/T	0.120 (T)	Intronic SNP	15	5	4	9	6	9	0.085
9:81589417	rs2777772	T/G	0.120 (G)	Intronic SNP	18	3	3	6	9	9	0.0003
9:81589329	rs2796464	T/C	0.120 (C)	Intronic SNP	19	3	4	9	6	9	0.02
9:81589690	rs2796465	G/A	0.120 (A)	Intronic SNP	19	3	2	8	7	9	1.93×10^{-6}

Table 22 SNPs found in intron 16/17 in a cohort of 24 CD patients and 24 healthy controls.

Results from intron 16/17 sequencing in Discovery cohort. Table shows chromosomal location, SNP ID numbers, type of variation, the number of patients and controls that are homozygous for the major or minor allele, the number of patients that are heterozygous and the p value for the difference in minor allele frequency between patients and controls. The Students T Test was used to generate a p value, heterozygous and homozygous (minor allele) individuals were grouped together i.e. the presence of one copy of the minor allele was used as the "risk factor". In total 8 SNPs and one deletion were found, of these variations three had statistically significant allele frequencies in CD patient's vs healthy controls: rs2777772 (p=0.0003), rs2796464 (p=0.0201), rs2796465 ($p=1.9337\times10^{-8}$).

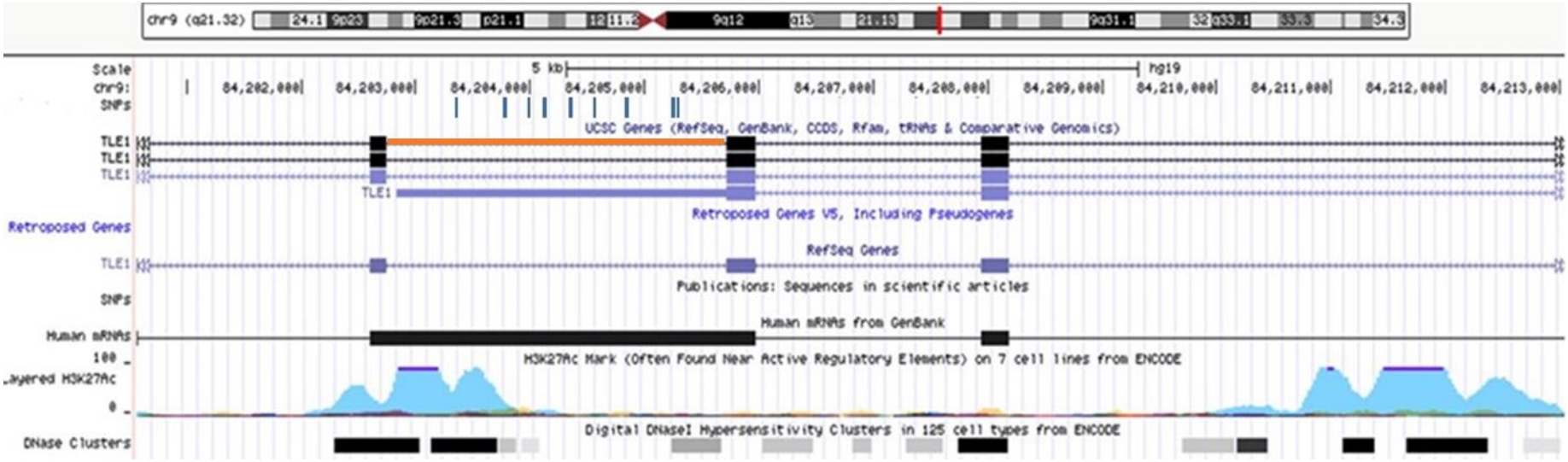


Figure 54 Features of potential functional significance in the TLE1 intron 16/17 region. Adapted from the UCSC genome browser

Figure showing functional SNPs of particular interest identified using SuRFr pipeline in top row and *TLE1* transcripts, histone acetylation and DNase 1 hypersensitivity clusters below, respectively. Intron 16/17 is shown in orange on the primary (top) *TLE1* transcript, SNPs in intron 16/17 that were identified by the SuRFr pipeline are depicted by blue lines above the first *TLE1* transcript.

6.4.2 Haploreg analysis of SNPs found in Discovery cohort

Haploreg is a tool designed to help prioritise variants from GWA studies (Ward & Kellis, 2012b). It provides data on chromatin state, evolutionary conservation and predicted transcription factor binding sites/regulatory motif alterations. The data provided is based on a combination of information from the ENCODE project, 1000 genomes project and genome browsers (UCSC, Ensembl). In this context, Haploreg was used as a tool to analyse whether any of the SNPs found in this Discovery cohort interrupted predicted transcription factor binding sites that have been implicated in IBD.

All sixteen SNPs found in the Discovery cohort were batch input into Haploreg for analysis, no data was available for rs372238712, rs376377585 and rs373720466. Table 25 shows the results from this analysis, in addition to regulatory motif alterations, MAF in African, Asian, American and European populations and dbSNP functional annotation are shown. Information from the ENCODE project/UCSC is used to produce promoter and enhancer histone marker information in addition to DNase 1 hypersensitivity and proteins bound. The only SNP to be located in a predicted promoter region according to histone markings was rs147523347, located in exon 2 of *TLE1*. Eight of the thirteen SNPs analysed were in DNAase1 hypersensitivity sites: rs147523347 rs7023704 (11 cell types) , rs2796471 (8 cell types) , rs6559629 (4 cell types), rs187303459, rs62578588, rs2796466 (1 cell type). Five proteins are predicted to bind to the rs147523347 region: POL2, CTCF, POL24H, and TAF. Eight of the thirteen SNPs analysed interrupt predicted regulatory motifs, 27 motifs are interrupted in total, of which three: PAX 5, RXRA and FOXJ2 are interrupted by two SNPs. Eight of the thirteen SNPs have MAF frequencies between 0.23 and 0.51 in European populations, frequencies in African , American and Asian populations are similar or lower. Two SNPs (rs147523347, rs7019824) have a MAF of 0.01, rs62578588 has a MAF of 0.08 in European populations and there is no allele frequency data available for rs114633202 and rs187303459. Huvec (human umbilical vein endothelial) cells have enhancer histone markings in the rs7023704 and rs2796471 SNP regions. Enhancer markings in HMEC (Human mammary endothelial cells) are seen in the

rs7019824 and rs6559629 SNP region. In the hepatocellular carcinoma cell line HEPG2 enhancer markers are seen in the rs911638, rs2777772 and rs2796464 SNP regions.

6.4.3 Analysis of SNPs in a European cohort

In total, fifteen SNPs were discovered in exons 2 and 6 and introns 15/16 and 16/17 in a Discovery cohort of 24 CD patients and 24 healthy controls. Of these 16 SNPs, 6 of the SNPs found in intron 15/16 were present at different frequencies in CD patients and controls.

A combined European dataset comprising 15,694 CD cases and 14,026 controls was analysed for the frequency of the 15 SNPs found in the Discovery Cohort. This dataset has been published by Franke et al, 2010. Data for six SNPs was available, none of which showed significant differences between CD cases and controls (Table 23). Data is shown for rs2796471 (p=0.15), rs6559629 (p=0.55), rs911638 (p=0.13), rs2777772 (p=0.43), rs2796464 (0.43) and rs2796465 (p=0.22).

The possible epistatic interaction between *NOD2* and *TLE1* has been investigated in this cohort previously by Dr Elaine Nimmo and *TLE1* has not been shown to alter the effect of *NOD2* mutations.

SNP	P Value
rs2796471	0.15
rs6559629	0.55
rs911638	0.13
rs2777772	0.25
rs2796464	0.43
rs2796465	0.22

Table 23 TLE1 SNPs found in Discovery cohort are not associated with CD in a larger European Replication cohort

TLE1 SNPs found in Discovery cohort were analysed in a larger European cohort (Franke et al., 2010) , p values shown are Bonferroni corrected.

6.4.4 Analysis of SNPs in Scottish cohort

To further investigate these findings, the frequency of these fifteen SNPs were analysed in a larger Scottish Replication cohort of 393 individuals: 203 CD patients and 190 healthy controls. This cohort consists of Scottish patients and control data collected after the Nimmo et al study in 2011 and there was no overlap between patients and controls between the two cohorts. This cohort has been genotyped by the WTCRF. All 15 SNPs were analysed in this larger cohort, six SNPs met quality control criteria: rs2796464, rs2777772, rs911638, rs2796465, rs62578588, rs2796471 (Table 24).The remaining SNPs either did not meet QC standards or had not been genotyped. The association study was conducted using PLINK v1.9 with the kind help of Dr Nick Kennedy (Purcell et al., 2007). Uncorrected p values for all 6 SNPs ranged from 0.65 to 0.96 and were therefore all non- significant. Odds rations ranged from 0.99 to 1.14.

The possible epistatic interaction between *NOD2* and *TLE1* has been investigated in this cohort previously by Dr Elaine Nimmo and *TLE1* has not been shown to alter the effect of *NOD2* mutations.

SNP	Minor Allele Frequency	Chi Squared	P value	Odds Ratio
rs2796464	0.26	0.0061	0.94	1.01
rs2777772	0.26	0.049	0.82	1.04
rs911638	0.26	0.049	0.82	1.04
rs2796465	0.26	0.038	0.85	1.04
rs62578588	0.077	0.208	0.65	1.14
rs2796471	0.37	0.0026	0.96	0.99

Table 24 Association study of SNPs found in Discovery cohort in larger Scottish replication cohort

Association analysis of SNPs discovered in Discovery cohort in a larger Scottish replication cohort of 203 CD patients and 190 healthy controls. Minor allele frequency, chi squared statistics, p values and odds ratios were calculated using PLINK. Uncorrected p values are shown.

Variant	Reference allele	Alternative allele	AFR freq	AMR freq	ASN freq	EUR freq	Promoter histone marks	Enhancer histone marks	DNase Hypersensitivity	Proteins bound	Motifs changed	dbSNP functional annotation
rs147523347	A	G	0	0	0	0.01	7 cell types		11 cell types	POL2,CTCF,POL24H, TAF,HEY1	GATA,GCNF	synonymous
rs114633202	A	G	0.02	0	0	0						synonymous
rs7023704	C	G	0.04	0.23	0.27	0.29		Huvec	11 cell types		EBF,RXRA	intronic
rs2796471	A	G	0.19	0.24	0.05	0.32		Huvec	8 cell types			intronic
rs6559629	G	A	0.49	0.49	0.6	0.51		HMEC	4 cell types		ERα,HIF1,MIZF, TATA,XBP-1	intronic
rs7019824	C	T	0.24	0.04	0.12	0.01		HMEC			Irf,NFE2,Nanog,Pax5,SP1,YY1,Zbtb3	intronic
rs187303459	C	T	0.01	0	0	0			GM12891		Irf	intronic
rs62578588	T	C	0.11	0.1	0.04	0.08			Th1			intronic
rs2796466	G	T	0.34	0.33	0.03	0.44			Osteoblasts		CTCF,Foxj2,Foxp1, HDAC2,HNF4,Pax5,RXRA,TATA,p300	intronic
rs911638	C	T	0.05	0.14	0	0.24		HepG2			Spz1,Zfp410	intronic
rs2777772	T	G	0.05	0.14	0	0.23		HepG2				intronic
rs2796464	T	C	0.04	0.15	0	0.24		HepG2				intronic
rs2796465	G	A	0.03	0.14	0	0.24					CEBPB,Foxj2	intronic

Table 25 Haploreg output

Haploreg output for 13 of the 16 SNPs discovered in the initial cohort. Data for rs372238712, rs376377585 and rs373720466 was not available. From left to right, column are: variant (rs ID), reference allele (major allele), alternative allele (minor allele), MAF in four populations (African, American, Asian and European) (all from dbSNP), promoter, enhancer markers (UCSC/Ensemble), DNase 1 hypersensitivity (UCSC/Ensemble), proteins bound (ENCODE), regulatory motif changes (ENCODE) and dbSNP functional annotation.

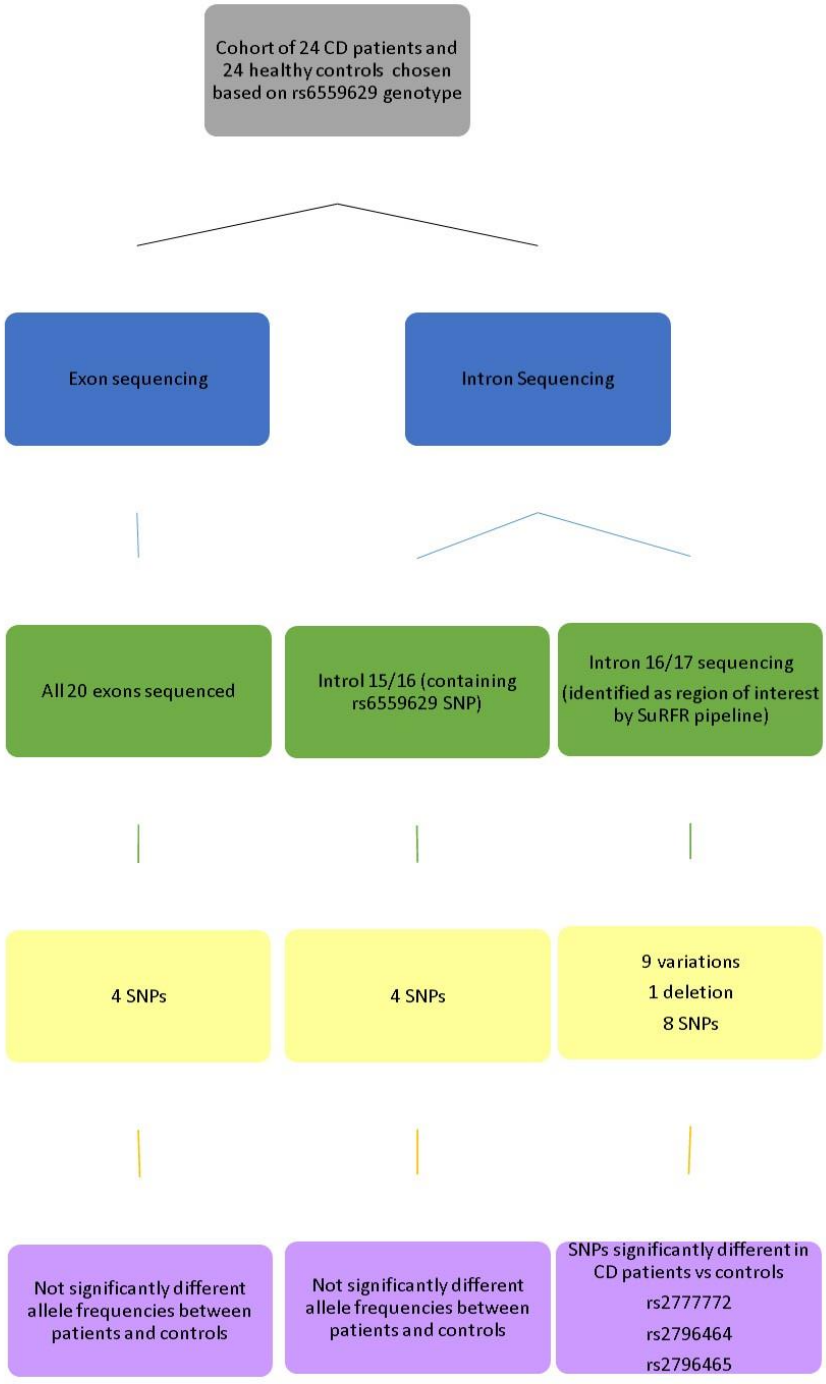


Figure 55 Summary of results from TLE1 sequencing

Summary figure showing regions of the *TLE1* sequenced in a cohort of 24 CD patients and 24 healthy controls. Results from the sequencing are summarised in the lower boxes.

6.5 Discussion

The aim of the work described in this chapter was to identify any mutations that may underlie the association between the rs655629 SNP in *TLE1* and Crohn's disease in Scottish populations (Nimmo et al., 2011) . A Scottish cohort of 24 ileal CD patients and 24 healthy controls were selected as an initial Discovery cohort. To enrich for mutations underlying the rs6559629 association CD patients with the rs6559629 risk allele and healthy controls without the risk allele were selected for this cohort.

All twenty exons of TLE1 were amplified and sequenced in the Discovery cohort, two SNPs were discovered: rs147523347 (p=1) and rs114633202 (p=0.56). Both SNPs encoded synonymous changes and the minor allele was present in both CD patients and healthy controls.

Additionally two *TLE1* introns: 15/16 and 16/17 were identified by association with CD and the SuRFr pipeline, respectively, as of regions of particular interest and were also sequenced in the Discovery cohort. Thirteen variants (including one deletion) were found in intronic regions of which rs2777772 (p=0.003), rs2796465 (p=0.00095) and rs2796464 (p=0.02) were present at different frequencies in patients and controls and achieved statistical significance. However, it is important to note that this significance does not reflect differences in the general population due to selection of controls and patients according to rs6559629 genotype.

Where possible, each of the 15 variants found in the Discovery cohort was followed up in a larger Scottish (CD=203, HC=190) and European replication cohort (CD=15,694, HC=14,026) of 203 CD patients and 190 healthy controls. None of the six SNPs were present at significantly different allele frequencies in CD patient's vs healthy controls in the replication cohort.

In this study a case control cohort were specifically selected based on rs6559629 genotype, therefore any statistical tests performed on the Discovery cohort are not a

true representation of population allele frequencies. This cohort allowed identification of variants in LD with the rs6559629 SNP that were not typed by the study conducted by Nimmo et al in 2011. This work has focused on the Scottish population, the rs6559629 association with CD has not been replicated in the UK (unpublished data, UK IBD consortium; (Franke et al., 2010)). This may be due to differences in ancestry/founder populations between Scottish and English populations leading to genetic heterogeneity. The Scottish population has been shown to have stronger Scandinavian influence than the English population which is reflected in differences in mitochondrial and Y chromosomal DNA (Capelli et al., 2003; Goodacre et al., 2005). This is complemented by geographic differences in *NOD2* genotype frequencies, which show that the *NOD2* 1007*fsInsC* and R702W variants are less common in the Scottish population than in the USA or central Europe (Arnott et al., 2004).

Figure 56 NOD2 allele frequency in Europe

Frequency of NOD2 L1007insCfs allele across Europe, frequencies are higher in central and southern Europe when compared to northern Europe.

Furthermore, the clear north-south divide in IBD incidence implies genetic differences between populations in addition to environmental differences. Future work could

involve deep sequencing of the *TLE1* region in both Scottish and Scandinavian populations. Recent studies have focused on deep sequencing of susceptibility genes discovered by GWA studies in order to identify low frequency, causative variants. In IBD, deep sequencing has revealed rare variants in loci known susceptibility loci including: *NOD2*, *IL23R* and *XBP1* (Balzola et al., 2012; Kaser et al., 2008).

Two synonymous exonic SNPs were found in the Discovery cohort used in this study: rs147523347 (exon 2), rs114633202 (exon 6). Both SNPs had a low MAF as would be expected of a rare causative variant (between 0.0009-0.003). However, neither showed different allele frequencies between CD patients and healthy controls: rs147523347 (p=1), rs114633202 (p=0.60). Data for these SNPs was not available in the much larger cohort of 393 (203 CD patients and 190 healthy controls). Although these SNPs are not present at statistically significant frequencies, follow up analyses i.e. deep sequencing of the *TLE1* region are warranted, particularly with regard to the rs114633202 SNP as it was not observed in CD patients and therefore may have a protective effect. Both these SNPs are synonymous, however they may still be of functional significance, it has been shown that synonymous variants in the Multi drug resistance gene (*MDR1)* have functional consequences. *MDR1* encodes an ATP dependant, membrane bound, drug efflux pump. Synonymous mutations in exon 26 of the *MDR1* gene encode a rare codon which delays co translational protein folding. This results in a protein product with altered structure, which interacts differently with both drug substrates and inhibitors (Kimchi-Sarfaty et al., 2007). As these are exonic SNPs, even synonymous variants could have a profound effect on translation efficiency, translation efficiency or other factors such as transcription factor binding.

In addition to the exonic SNPs discussed there were a number of promising intronic SNPs identified in the Discovery cohort. The indel variant, rs376377585 was seen as a seven base pair deletion in three healthy controls. This variant may have a protective effect and warrants further investigation in larger cohorts as data for this variant was not available in the larger European and Scottish cohorts used in this study. Other intronic variants of interest include those in intron 16/17 which were present at

statistically significant allele frequencies between cases and controls, although they did not show significance in the larger cohorts discussed in this study it may be of interest to look at the frequency of these variants in Scandinavian cohorts.

A key limitation of this work is that it may have lacked the power necessary to detect statistically significant differences in allele frequencies. Due to the study design, the exact power cannot be calculated as the cohort was enriched for CD cases with the rs6559629 risk allele. However, if this was a standard GWA study the number of people needed to achieve 80% power to replicate the association between rs6559629 and CD would be approximately 10,000 cases and controls. This is not possible in a Scottish cohort, however the UK replication cohort used met these criteria. Given that the rs6559629 SNP was not associated with CD in this cohort, it may be the case that the Nimmo et al finding was an artefact or that the association is a Scottish specific finding, as previously discussed. Additionally, the strongest association between *TLE1* and CD was seen in ileal CD patients, the European replication cohort did not have sub phenotypic data and therefore the *TLE1* association may have been masked. Current projects in our laboratory involve the collection of DNA samples from Scottish CD patients as well as Scandinavian and other European CD patients, analysis of *TLE1* in these patients will help to determine the significance of the association between rs6559629 and ileal CD in the found by Nimmo et al in the Scottish population.

7 Discussion

The work presented in this thesis investigates the role of the *TLE1* gene in Crohn's disease. *TLE1* was identified as a potential CD susceptibility gene by Nimmo *et al* (2011). A yeast 2 hybrid study showed that TLE1 interacts with the known CD susceptibility protein NOD2. Additionally an association study showed that polymorphisms in *TLE1* were associated with CD in a Scottish cohort (healthy controls n=1515, CD = 648). The strongest association was observed between the rs6559629 SNP in intron 15/16 of *TLE1* and cases with ileal CD ($p=3.1 \times 10^{-5}$). Furthermore, the rs6559629 risk allele was shown to increase the effect of polymorphisms in *NOD2*, however this it is important to note that only a relatively small proportion of the cohort had *NOD2* mutations (22.7%).

It is important to note that neither the association between SNPs in *TLE1* nor the finding that *TLE1* polymorphisms increase the effect of *NOD2* mutations found in CD have been replicated in other cohorts. For example, data from a larger UK cohort of 4307 healthy controls and 1903 CD cases genotyped for the three *NOD2* risk alleles and the rs6559629 SNP by the Wellcome Trust Clinical Case Consortium did not replicate either of these findings. Although it appears the *TLE1* polymorphisms may not be important in CD pathogenesis in different populations, they may still be valid in Scottish populations as a result of different ancestral lineage of the Scottish population compared to the English population. It is also important to note that although TLE1 has not been identified by GWA studies in other UK cohorts, interacting partners of TLE1 have been identified such as *NOD2* (Jostins et al., 2012).

My work, analyses *TLE1* sequence, TLE1 expression and interacting partners of TLE1 in the context of Crohn's disease. In this chapter, the key findings are discussed alongside limitations and further work needed to determine whether TLE1 plays a role in IBD pathogenesis.

7.1 TLE1 expression and localisation

7.1.1 Key findings

Expression and localisation of TLE1 was analysed by immunohistochemistry, in ileal resection samples from healthy controls and CD cases. CD cases were grouped in two according to whether they had a NOD2 mutation. TLE1 expression was detected in the nucleus and the cytoplasm. Overall, expression in CD cases varied from 70% + positivity to less than 20% positivity. These differences did not correspond to NOD2 genotype. However, it is important to note that the three NOD2 mutations: R702W, G908R and L1007fs were grouped together in this study. Due to the low frequency of these mutations in the Scottish population, ethical concerns did not allow analysis of each genotype individually. This may have been the cause of the range of expression seen in the NOD2 mutant group.

7.1.2 Future work: TLE1 and NOD2 function in Paneth cells

TLE1 and NOD2 are expressed in Paneth cells of the small intestine. This expression is seen in both the nucleus and the cytoplasm of Paneth cells. As previously discussed, Paneth cells are located at the base of the small intestinal crypts and secrete antimicrobial peptides in order to protect the crypts from bacterial invasion. The autophagy protein ATG16L1, the bacterial sensor NOD2 and numerous UPR genes have been implicated in IBD pathogenesis. It has previously been postulated that autophagy, UPR defects and defects in bacterial recognition converge to cause Paneth cell dysfunction seen in IBD (Figure 57). Both ATG16L1 and NOD2 variants are associated with Paneth cell dysfunction. Interestingly they appear to have cumulative effect i.e. cases with risk alleles in both genes had increased numbers of abnormal Paneth cells when compared to those with one risk allele (Vandussen et al., 2014). Increased expression of the ER stress markers GRP78 and EIF2α (phosphorylated) has been shown in cases with the ATG16L1 T300A risk allele. This was shown by immunohistochemical analysis of protein expression (Deuring et al., 2014).

Results from this thesis suggest that *TLE1* may be involved in the unfolded protein response. This is shown by differential expression of *TUSC3* following *TLE1*

knockdown. *TUSC3* has been implicated in the UPR; *TUSC3* knockdown leads to a decrease in expression of both the UPR sensor, BiP and the downstream UPR transcriptional regulator, CHOP (Horak et al., 2014). *TLE1* appears to act as a transcriptional co repressor for *TUSC3*; upon *TLE1* knockdown, *TUSC3* expression increases. The role of *TUSC3* has not been shown in the gut, future work may involve analysing whether *TLE1* influences *TUSC3* expression directly and whether the TLE1 and TUSC3 proteins interact.

The NAD dependant deacetylase, SIRTUIN (SIRT1) has also been shown to interact with TLE1. TLE1 was identified as a potential interacting partner of SIRT1 using a yeast two hybrid assay; these results were confirmed in HeLa cells overexpressing TLE1 and SIRT1 (Ghosh et al., 2007b). SIRT1 has been implicated in both the ER stress response and autophagy. These are both processes that are dysregulated in IBD and are important to Paneth cell function. In mouse models of obesity immunohistochemical analysis of liver tissue shows increased SIRT1 expression leads to decreased expression of ER stress markers: CHOP, GRP79. Additionally SIRT1 overexpression leads to a loss of phosphorylated EIF2α, as shown by western blotting for phosphorylated and unphosphorylated EIF2α isoforms (Li et al., 2011). Reduced expression of SIRT1 has also been shown in the colonic mucosa of IBD cases, particularly in the intestinal crypts, where both NOD2 and TLE1 are expressed. SIRT1 expression is principally nuclear in the colonic mucosa, however there are also very low levels of cytoplasmic staining in the crypts (Melhem et al., 2015). SIRT1 over expression appears to stimulate autophagy in the intestinal epithelial cell line, HCT116. Co immunoprecipitation experiments using HeLa cells show that SIRT1 interacts with the autophagy proteins: ATG 6, 7 and 8 (I. H. Lee et al., 2008). Results from this thesis suggest that *TLE1* does not influence expression of SIRT1, however the interaction between TLE1 and SIRT1 may provide a link between TLE1, the UPR and autophagy.

TLE1 knockdown altered expression of two predicted mTOR interactors: RIOK1 and CCND1. mTOR is a serine threonine kinase which is involved in numerous metabolic

pathways e.g. protein translation, ribosome biogenesis and cell growth. It phosphorylates over 800 proteins. There are numerous lines of evidence suggesting mTOR is involved in autophagy. Under normal conditions (non- starvation) mTORC1 phosphorylates transcription factor EB (TFEB). Phosphorylated TFEB is retained in the cytoplasm. However if mTOR is inhibited, for example by nutrient starvation, TFEB translocates into the nucleus (Martina, Chen, Gucek, & Puertollano, 2012). Activated, unphosphorylated TFEB was shown to initiate transcription of 11 out of 51 autophagy related genes, these included: *UVRAG,ATG9B* and *WIPI* (Settembre et al., 2011). mTOR inhibition has also been shown to initiate autophagy in model organisms (Ravikumar et al., 2004). mTORC1 forms a complex with the autophagy proteins ULK1 and ATG13 in HEK 293T cells (Hosokawa et al., 2009).

Taken together, it could be hypothesised that TLE1 may contribute to the Paneth cell dysfunction seen in IBD via interaction with/regulation of UPR (SIRT1 and TUSC3) and autophagy regulators (SIRT1). Further investigation of this by analysing colocalisation and of NOD2 and TLE1 along with autophagy and UPR proteins may help to elucidate this. The use of animal models to study the effects of TLE1 and NOD2 knockdown on UPR and autophagy gene expression in Paneth cells may also be of interest.

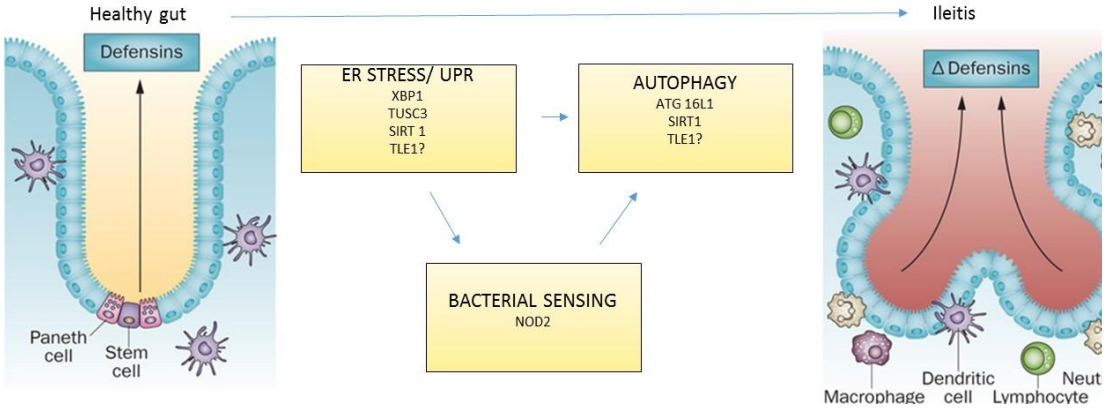


Figure 57 The UPR, bacterial sensing and autophagy contribute to Paneth cell dysfunction

Adapted from (Van Limbergen, Radford-Smith, & Satsangi, 2014). Schematic diagram showing development of the ileitis from normal intestinal crypts. Autophagy, bacterial sensing and the ER stress response are all thought to play a role in the development of ileitis. These three biological pathways have also been shown to interact.

7.2 Sequence analysis of *TLE1*

7.2.1 Key findings

The strategy used in this study was different to standard GWAS studies in that the cohort was specifically selected according to rs6559629 genotype. Additionally only CD cases with ileal CD were studied as this group of cases showed the strongest association with rs6559629. This Discovery cohort allowed identification of variants in LD with the rs6559629 SNP that were not typed by the study conducted by Nimmo et al in 2011. A Discovery cohort of 24 CD cases and 24 ileal CD cases with the rs6559629 risk allele and alternative allele, respectively were chosen for this study. Due to this enrichment p values shown for the Discovery cohort are not representative for the general Scottish population and are only provided as a guideline to suggest differences in frequency. These frequency of these variants was then analysed in two replication cohorts: a Scottish cohort (203 CD cases, 190 controls) and a combined European cohort (6,333 CD cases, 15,056 controls). None of the variants identified were present at statistically different frequencies between CD cases and controls.

The finding suggesting *TLE1* is not associated with CD in larger European cohort poses the key question: is the association shown in the Nimmo *et al* study an artefact of the relatively small numbers used or is it representative of a population specific association? Differences between Scottish and English ancestry are discussed in depth elsewhere. In summary, the Scottish population has been shown to have stronger Scandinavian influence than the English population which is reflected in differences in mitochondrial and Y chromosomal DNA (Capelli et al., 2003; Goodacre et al., 2005). This is complemented by geographic differences in *NOD2* genotype frequencies, which show that the *NOD2* 1007*fsInsC* and R702W variants are less common in the Scottish population than in the USA, UK or Europe and are almost absent in Japan, Korea and China (Arnott et al., 2004). Furthermore, the clear north-south divide in IBD incidence implies genetic differences between populations in addition to environmental differences. Low *NOD2* mutation frequency in Scottish and Scandinavian cohorts may be a reflection of Viking ancestry compared to more prevalent Anglo-Saxon ancestry in England. Population specific associations have

been identified in other IBD susceptibility loci, for example insertion/deletion variants in the IGRM locus that have been associated with CD in European populations, these are not found in Japanese populations (Prescott et al., 2010).

Although SNPs found in *TLE1* were followed up in the largest available Scottish cohort, this cohort may still have been underpowered to detect significant differences in carriage frequency. These SNPs were also followed up in a larger combined cohort of cases and controls across Europe, results did not show any differences in carriage frequency. Current projects in our lab involve collection of samples from both Scottish and European cohorts, this will enable us to determine whether *TLE1* is associated with CD in larger Scottish cohorts and Scandinavian cohorts.

7.2.2 Future work: Deep Sequencing of the *TLE1* locus

Although the size of the Discovery cohort allowed detection of mutations at approximately 1% frequency in the population, it may have been underpowered to detect more subtle differences in the frequency of the 15 variants found in the Discovery cohort. The exact power of the Discovery cohort cannot be calculated due to study design. Additionally the size and approach used may not have allowed identification of rare private mutations. Furthermore, Sanger sequencing was used to sequence these samples, this type of sequence only produces a single read from each sample. The use of technology to get multiple reads would allow identification of mutations present in cell subtypes present at less than 1% frequency in the sample.

Deep sequencing involves generating thousands of reads for the same sequence which allows for increased coverage and identification of rare mutations that may only be present in a very small proportion of the cells from which the DNA sample originates. An alternative approach to identifying variants underlying associations identified by GWAS studies involves deep sequencing of genes in larger groups of cases. More recently this has become more accessible due to the decrease in costs of next generation sequencing technologies.

A recent study deep sequenced 50 known IBD susceptibility genes in a cohort of 350 cases and controls, mutations found were then genotyped in a much larger cohort of 16,054 Crohn's disease cases, 12,153 ulcerative colitis cases and 17,575 healthy controls. Private causative and protective variants in *NOD2* as well as other IBD susceptibility genes were identified. This study used a mixed European cohort which would not be appropriate for deep sequencing *TLE1* as the association between CD and *TLE1* have not been observed outside of the Scottish population, however, a study similar to the initial cohort would be feasible in the Scottish population.

7.3 TLE1 interacting proteins

7.3.1 Key findings: TLE1 and XBP1

Analysis of the rs6559629 SNP in TLE1 revealed it interrupts a potential XBP1 binding site. In this work a ChIP assay was designed optimised and implemented for XBP1 and results from cell lines show that exogenous XBP1 does not bind to its predicted binding site in *TLE1*.

7.3.2 Future Work: TLE1 and XBP1

The rs6559629 site is not in a known promoter region, the nearest TSS is the *TLE1* TSS which is approximately 10Kb downstream of the XBP1 recognition motif in intron 15/16 of *TLE1*. Many studies have focused on the binding of XBP1 to promoter regions for example XBP1 has been shown to bind to promoter regions of GPR43, CHOP and BECLIN1 using ChIP assays (Ang et al., 2015; Margariti et al., 2013; Shao et al., 2015). Although the potential binding site interrupted by rs6559629 was not close to a TSS it was targeted by this study as it had sequence overlap for two core XBP1 binding motifs. Additionally although Acosta et al suggest that many of XBP1 binding motifs occur at transcriptional start sites, their work used proximal promoter arrays and hence a long range transcriptional activation mechanism for XBP1 cannot be ruled out. Many transcriptional regulators have been shown to have long range effects in addition to short range ones. For example, the transcriptional activator, Estrogen receptor (ER) has only 22% of its targets near TSS, the remaining targets are

long range targets, some of which are up to 100Kb away from a TSS. Furthermore these distal ER binding sites are functional and activate transcription of downstream genes (Carroll et al., 2006). Although according to this work XBP1 does not bind the rs6559629 site, given that *XBP1* is an IBD susceptibility gene in its own right, it would be of future interest to explore long range effects of XBP1 to further elucidate its role in IBD pathogenesis.

In this work binding of both the spliced and unspliced forms of XBP1 was analysed. Recent work has shown that the XBP1 (us) is a functional interacting protein. It has been shown to interact with the autophagy protein FOXO1, by co-immunoprecipitation and colocalisation was shown using immunocytochemistry. It was also suggested that XBP1(us) facilitates degradation of FoxO1 by the proteasome (Zhao et al., 2013). Although the work presented suggests XBP1 (us) does bind to the same positive control regions as XBP1 spliced, it is possible that XBP1 (us) has different targets and mechanisms of action to XBP1 (s). As the positive control used in this work only worked for XBP1 (s) it could be the case that the ChIP assay did not accurately detect the XBP1 (us) targets. The use of high throughput technologies such as chip on chip to identify XBP1 (us) targets would be of interest, to identify whether XBP1 (us) binds to the rs6559629 site as well as to compare XBP1 (s) and XBP1 (us) functionality.

7.3.3 Key Findings: TLE1 and NOD2

In 2011, the NOD2/TLE1 interaction was identified in HEK293 cells over expressing NOD2 and TLE1. As part of this project attempt was made to further characterise this interaction in cell lines using western blotting and co immunoprecipitation however a functional antibody could not be identified.

7.3.4 Future Work: TLE1 and NOD2

Further characterisation of this interaction to identify which regions of the proteins are involved in this interaction was attempted. However, this was limited by time constraints and a lack of a functioning TLE1 and NOD2 antibody. Further characterisation using the recently identified NOD2 antibody is warranted.

Additionally, the direct interaction between TLE1/NOD2 has not yet been confirmed in human tissue samples. This is necessary to confirm that results from cell lines are representative of human cells. Currently this is not possible due to the fact that both the functional NOD2 and TLE1 antibodies available are raised in the same species and cannot, therefore, be used in co localisation experiments such as immunohistochemistry using fluorescently tagged markers.

7.4 Future Perspectives

7.4.1 Analysing methylation of TLE1

This work presented in this thesis did not analyse either epigenetic changes in *TLE1*. Alterations in methylation of loci have been shown to alter transcription factor binding and gene expression.

The rs6559629 SNP creates or destroys a potential CpG site by substitution of A to G. The methylation status of this CpG site may influence XBP1 binding. As previously discussed an appropriate pyrosequencing assay could not be designed for this site. Future work could use chip technology to analyse global methylation in Scottish adults in the rs6559629 region. The rs11190140 SNP in NKX23 was shown to be associated with IBD in a Pennsylvanian cohort and it also interrupted CpG site. Bisulphite sequencing was used to analyse the methylation status of the rs11190140 site and showed the C allele was methylated in a subset of IBD cases with the C allele. Furthermore, methylation of the site was shown to alter binding of the transcriptional

regulator NFAT1 in B cells using a ChIP assay (John et al., 2011). As the methylation status of the rs6559629 site is not known in the cell lines used it may not accurately reflect the situation in IBD cases. Previous work has analysed global methylation in paediatric IBD cases vs controls. Four sites in *TLE1* are included as part of these studies, these sites did not include the rs6559629 SNP and were not differentially methylated between cases and controls and hence the methylation of the rs6559629 site warrants further investigation (Adams et al., 2014).

TLE1 promoter methylation is altered in hematologic malignancies. The CpG island located in the *TLE1* promoter region (approximately 500bp from the TSS) was analysed by Fraga *et al* using a combination of bisulphite sequencing and methylation specific PCR. It has been shown that the *TLE1* promoter is only methylated in haematological cell lines; analysis of intestinal epithelial cell lines such at HCT 116 and SW 480 are not methylated in this region (Mario F Fraga et al., 2008b). The rest of the *TLE1* gene has not been analysed for methylation with regard to disease status. SNPs in the promoter region of *TLE1* were not analysed as part of the Nimmo *et al* study. Recent work in our lab had analysed genome wide methylation in paediatric cases and controls. This work led to the identification of mir21 as differentially methylated between cases and controls. Furthermore, hypomethylation of the region, confirmed by an Illumina methylation chip and subsequent pyrosequencing, corresponded to increased *VMP1/mir21* expression in leukocytes and inflamed intestinal tissue of IBD cases (Adams et al., 2014; Noble et al., 2010). A Scottish cohort was used in this study and was shown to be sufficiently powered to detect changes in methylation, a similar cohort could be used to investigate the methylation of the *TLE1* promoter and rs655929 region. It is important to note that *TLE1* is differentially methylated according to sex and this would need to be corrected for in a patient cohort (J. Liu, Morgan, Hutchison, & Calhoun, 2010).

7.4.2 Summary of future work

This thesis has raised new questions as to how *TLE1* may be involved in IBD pathogenesis. Key questions that should be the focus of future work are listed below.

1. Deep sequencing of the *TLE1* locus in a Scottish cohort.

   The work described in this thesis shows sequence data for *TLE1* in a cohort of Scottish cases. This was conducted in order to identify mutations underlying the rs6559629 association with ileal CD. The study design used a small cohort enriched with CD cases enriched for the rs6559629 risk allele. Although some variants were identified, none were limited to cases or healthy controls. Recent studies have used cohort sizes similar to our Scottish cohort (approximately 393 cases and 197 controls) to deep sequence known IBD susceptibility genes and identify rare variants underlying associations from GWAS studies.

2. Role of TLE1 in Paneth cells

   This thesis shows that TLE1 is expressed in the Paneth cells of the small intestine. Colocalisation of TLE1 and NOD2 in Paneth cells has not been shown as both antibodies are raised in the same species. The use of serial sections to analyse TLE1 and NOD2 staining in the same cells is currently being undertaken in our laboratory. In recent years, there have been great advances in microscopy techniques such as super resolution photo activated localization microscopy (PALM). These techniques allow for live cell imaging and deep tissue imaging would be of interest to study NOD2 and TLE1 expression and localisation in the human gut. Although these types of microscopy are not yet commercially available, independent laboratories have built the microscopes in question.

3. TLE1 in autophagy

TLE1 knockdown studies showed that *TLE1* may be involved in autophagy and two potential *TLE1* targets are predicted to interact with the known autophagy protein, mTOR. Further characterisation of potential TLE1/mTOR interactions and the mechanism by which TLE1 would be of interest to future work as autophagy is dysregulated in IBD.

4. TLE1 in the UPR

   Although XBP1 (sp) does not appear to bind *TLE1*, this work shows that *TLE1* knock down alters expression of the UPR gene, *TUSC3*. The potential XBP1 (we) /TLE1 interaction and the mechanism behind *TUSC3*/*TLE1* interplay may provide additional mechanisms by which TLE1 is involved in IBD pathogenesis.

5. TLE1 in S1P signaling

   SGPL1 is a member of the S1pP signaling pathway and its expression is altered by *TLE1* knockdown. This work was done at the RNA level, future work could look at analysing this relationship at the protein level in addition to investigating the mechanism by which *TLE1* may be involved in S1P signaling

6. Methylation analysis of *TLE1*

   Methylation changes in *TLE1* have been associated with disease (M. F. Fraga et al., 2008). Work in our lab has shown differential methylation of genes between IBD cases and healthy controls in a Scottish cohort (Adams et al., 2014). The methylation status of the *TLE1* promoter and the rs6559629 region was not investigated by this study and hence would be of interest in future work. Methylation analysis could be analysed by genome wide analysis or individual CpGs could be analysed by pyrosequencing.

1 Appendix 1: Primers

TBP	TGCCCGAAACGCCGAATATA TTTCTTGCTGCCAGTCTGGA
SDHA	AGGGCATCTGCTAAAGTTTCAGA GATTCCTCCCTGTGCTGCAA
B ACTIN	GCCAGCTCACCATGGATGAT AATCCTTCTGACCCATGCCC
GAPDH	CGACCACTTTGTCAACTCA AGGGAGATTCAGTGTGGTG
UBC	GATTTGGGTCGCAGTTCTTG TCCAGCAAAGATCAGCCTCT

Table 26 Reference gene RT qPCR primers

RIOK1	GGGACTGGGATGAAGGAGTTG TGTCGATTTGCCTGTGGGTT
TUSC3	ATCCTTATAGTGATCTGGACTTTGA GCAAATCCCACTTGGCTTCAT
SGPL1	GGCGTGAGGAGAGTCTGAAA GCTCAAAGGCCTTCAACATCAG
CCND1	AGTGGAAACCATCCGCCG TCTGTTCCTCGCAGACCTCCA

Table 27 *RIOK1*, *TUSC3*, *SGPL1* and *CCND1* RT qPCR primers

XBP1 promoter 1	CTCGGCTAGAAGACACCTCC GTCCCAACTCCAGAGCCTC
XBP1 promoter 2	CAGGAGGCTCTGGAGTTGG TGGTAGGGAGGGAACAGGAT
XBP1 promoter 3	TTTCCCATCCTGTTCCCTCC GTAATTATTGCTTCCTCCTCTGC
XBP1 promoter 4	AACCTGGAATAGTGCTGCCC TGGCCAGGTGTTTGTGAAAG
XBP1 promoter 5	GCCCCTCTTTCACAAACACC CCTCGGCCTCCCAAAGTG
XBP1 promoter 6	GCCTGTAATCCCAGCACTTT TCACTGCAACCTTCGCCT AGGCGAAGGTTGCAGTGA GTTTGTACTATGAGCGTGCATT
XBP1 promoter 7	GGTATTCCCCTCCAACCCAG AGGTTTCAGTGAGCCGAGAT
XBP1 promoter 8	ATCTCGGCTCACTGAAACCT CCATCCGGAGTGACAGAATT
XBP1 promoter 9	TCAGTCTGGAAAGCTCTCGG CTCGGCGTCCATTGGTCC
XBP1 promoter 10	CATAGCCACGGTCCTGAAAC AGCAGAACTTTAGGGGTCCC

Table 28 *XBP1* promoter RT qPCR primers used for ChIP

Rs6559629 site primer set 1	TGCACAAGGGCAATTTCTCC GGAGAGCCTCTGTAAACCACT
Rs6559629 site primer set 2	AGTGGTTTACAGAGGCTCTCC AACCAGCTCTCTCAGGTCTC
Rs6559629 site primer set 3	GGAGACCTGAGAGAGCTGG AGAGATCCGAACATGTCCCA
Rs6559629 site primer set 4	TGGGACATGTTCGGATCTCT GGTGAAACCCCGTCTCTACT
Rs6559629 site primer set 5	GTTTCACCATGTTAGCCAGAATG AGGCAGACACCACACCAG

Table 29 rs6559629 site RT qPCR primers used for ChIP

TLE1 Exon number	Primer Sequence		Optimum annealing temp (°C)	PCR product length (bp)
	Forward	Reverse		
Exon 1	AACCTCGCTCCCCTTTG	AAATTAAGAGTTTCGGCCC	59.3	819
Exon 2	TCGTTTTCAGCCACTTCTAAC	GAGACTCCACACGCCAC	59.3	400
Exon 3/4	GAAGTGCAAATATGTGAGGC	CAAGAAGCAACCCTTGATG	61	400
Exon 5	CCATTTTATACACATTTCCAGAG	CCATTTCTACAGTGTGCCC	61	360
Exon 6	CATTCTCAGTTTTTCTTTCCC	CACACACGTAAAGCCATC	55.9	335
Exon 7/8	CATGTCTGTTTTCCCCTG	CCTCTTATCTCATTGTTTGCTC	61	423
Exon 9	ATTGCAAAGTCAGGGTCTC	CCGGCAGAAGCATAATTTAC	61	497
Exon 10	CAAACTGATAGAGAGCCTGC	AGAAGTTGCAAGAGGTCCC	61	313
Exon 11	GTACAACCCCAACCCG	CTCAAGCAGCAAAACCC	55	374
Exon 12	GAAAATTGGCTGAGGGAG	CGTACATTTCATATTTGTAGGGC	55.9	370
Exon 13	CCCTAGAAGCCTTACCCTTTC	AGACTGGGCGTCTTTGTC	56.7	455
Exon 14	GGGGATCTGTGAGTTTCTAA	GACACCAGAAATCTCCTTTGG	55.9	318
Exon 15	TTTGCACCAGCAGTAAATG	CAGGAAACAGAAGGGGC	55	538
Exon 16	CCATGAAACACCACTCTTAACTC	CATGCCTGGCATTCAAG	55.9	474
Exon 17	CACGGTCCAAAACTAACAAC	CAGCCCTGATCTCGTTTGTT	55.9	350
Exon 18	AGTCTGGACTTGGCTGATTAC	TATTGCATCCAGCTGCTC	55.9	331
Exon 19	GTCCTTAGGGCATGATAGTTC	TAGACTGTAGCCTTCTTGTCCC	56.7	431
Exon 20	GTCCTTAGGGCATGATAGTTC	TAGACTGTAGCCTTCTTGTCCC	56.7	431

Table 30 Primer details for *TLE1* exon sequencing

PCR/ sequencing primer sequences, optimal annealing temperature and PCR product length for each of the *TLE1* exons

PCR	GGAGAATTGCTTGAACCCGG CAGAAGCACGGTAGCCTTATCC
Sequencing	GGAGAATTGCTTGAACCCGG AGAAACGAGGTCTCACTGTGT CAGAAGCACGGTAGCCTTATCC AGGAGGTGGATGTGAAAGATCC

Table 31 Intron 16/17 PCR and sequencing primers

PCR	TACTGCAGACGGTCAGATGC GTTGTGCAGATCCCACACAG
Sequencing	TACTGCAGACGGTCAGATGC CTGCAGCATGAGAAGCAGAC CGTTACTGCAGACGGTCAGA CTGCAGCATGAGAAGCAGAC TACTGCAGACGGTCAGATGC GTTGTGCAGATCCCACACAG

Table 32 Intron 14/15 PCR and sequencing primers

2 Appendix 2: Patient Information

	Average Age		Gender	
	Male	Females	Males	Females
CD cases without *NOD2* variants	30	29	3	7
CD cases with *NOD2* variants	30	34	3	7

Table 33 Patient information for IHC study

	Average Age		Gender	
	Male	Females	Males	Females
Healthy controls	54	54	10	14
CD cases	66	64	12	12

Table 34 Age and gender information for CD cases and controls in Discovery cohort

References

Acosta-Alvear, D., Zhou, Y., Blais, A., Tsikitis, M., Lents, N. H., Arias, C., Dynlacht, B. D. (2007). XBP1 controls diverse cell type- and condition-specific transcriptional regulatory networks. *Molecular Cell*, *27*(1), 53–66. doi:10.1016/j.molcel.2007.06.011

Adams, A. T., Kennedy, N. A., Hansen, R., Ventham, N. T., O'Leary, K. R., Drummond, H. E., … Satsangi, J. (2014). Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn's Disease Implicates Epigenetic Alterations at the VMP1/MIR21 and HLA Loci. *Inflammatory Bowel Diseases*, *20*(10).

Aldhous, M. C., & Satsangi, J. (2010). The impact of smoking in Crohn's disease: no smoke without fire. *Frontline Gastroenterology*, *1*(3), 156–164. doi:10.1136/fg.2010.001487

Aldhous, M. C., Soo, K., Stark, L. a, Ulanicka, A. a, Easterbrook, J. E., Dunlop, M. G., & Satsangi, J. (2011). Cigarette smoke extract (CSE) delays NOD2 expression and affects NOD2/RIPK2 interactions in intestinal epithelial cells. *PloS One*, *6*(9), e24715. doi:10.1371/journal.pone.0024715

Ali, S. a, Zaidi, S. K., Dobson, J. R., Shakoori, A. R., Lian, J. B., Stein, J. L., … Stein, G. S. (2010). Transcriptional corepressor TLE1 functions with Runx2 in epigenetic repression of ribosomal RNA genes. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(9), 4165–9. doi:10.1073/pnas.1000620107

Allen, T., van Tuyl, M., Iyengar, P., Jothy, S., Post, M., Tsao, M.-S., & Lobe, C. G. (2006). Grg1 acts as a lung-specific oncogene in a transgenic mouse model. *Cancer Research*, *66*(3), 1294–301. doi:10.1158/0008-5472.CAN-05-1634

Andoh, A., Zhang, Z., Inatomi, O., Fujino, S., Deguchi, Y., Araki, Y., … Fujiyama, Y. (2005). Interleukin-22, a Member of the IL-10 Subfamily, Induces Inflammatory Responses in Colonic Subepithelial Myofibroblasts. *Gastroenterology*, *129*(3), 969–984. doi:http://dx.doi.org/10.1053/j.gastro.2005.06.071

Ang, Z., Er, J. Z., & Ding, J. L. (2015). is transcriptionally regulated by XBP1 in, 1–9. doi:10.1038/srep08134

Arnott, I. D. R., Nimmo, E. R., Drummond, H. E., Fennell, J., Smith, B. R. K., MacKinlay, E., … Satsangi, J. (2004). NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? *Genes and Immunity*, *5*(5), 417–25. doi:10.1038/sj.gene.6364111

Auphan, Nathalie, Didonato, Joseph, Caridad, Rosette, Helmburgh, A. (1995). Immunosuppression by glucocorticosteroids: inhibition of NF-κB activity though induction of IkB synthesis. *Science*, *270*, 286–289.

Balzola, F., Bernstein, C., Ho, G. T., & Russell, R. K. (2012). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease: Commentary. *Inflammatory Bowel Disease Monitor*, *12*(11), 126–127. doi:10.1038/ng.952

Barnich, N., Aguirre, J. E., Reinecker, H. C., Xavier, R., & Podolsky, D. K. (2005). Membrane recruitment of NOD2 in intestinal epithelial cells is essential for nuclear factor-??B activation in muramyl dipeptide recognition. *Journal of Cell Biology*, *170*(1), 21–26. doi:10.1083/jcb.200502153

Barnich, N., Hisamatsu, T., Aguirre, J. E., Xavier, R., Reinecker, H.-C., & Podolsky, D. K. (2005). GRIM-19 interacts with nucleotide oligomerization domain 2 and serves as downstream effector of anti-bacterial function in intestinal epithelial cells. *The Journal of Biological Chemistry*, *280*(19), 19021–6. doi:10.1074/jbc.M413776200

Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., Daly, M. J. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature Genetics*, *40*(8), 955–962. doi:10.1038/ng.175

Berndt, S. I., Skibola, C. F., Joseph, V., Camp, N. J., Nieters, A., Wang, Z., … Slager, S. L. (2013). Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nature Genetics*, *45*(8), 868–76. doi:10.1038/ng.2652

Boerries, M., Herr, R., Brummer, T., & Busch, H. (2015). Global gene expression profiling analysis reveals reduction of stemness after B-RAF inhibition in colorectal cancer cell lines. *Genomics Data*. doi:10.1016/j.gdata.2015.04.015

Bonen, D. K., & Cho, J. H. (2003). The genetics of inflammatory bowel disease. *Gastroenterology*, *124*(2), 521–536. doi:http://dx.doi.org/10.1053/gast.2003.50045

Bønnelykke, K., Sleiman, P., Nielsen, K., Kreiner-Møller, E., Mercader, J. M., Belgrave, D., … Bisgaard, H. (2014). A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nature Genetics*, *46*(1), 51–5. doi:10.1038/ng.2830

Boyle, J. P., Parkhouse, R., & Monie, T. P. (2014). Insights into the molecular basis of the NOD2 signalling pathway.

Brandon, N. J., & Sawa, A. (2011). Linking neurodevelopmental and synaptic theories of mental illness through DISC1. *Nature Reviews Neuroscience*, *12*(12), 707–722. doi:10.1038/nrn3120

Brighenti, E., Calabrese, C., Liguori, G., Giannone, F. a, Trerè, D., Montanaro, L., & Derenzini, M. (2014). Interleukin 6 downregulates p53 expression and activity

by stimulating ribosome biogenesis: a new pathway connecting inflammation to cancer. *Oncogene*, (November 2013), 1–11. doi:10.1038/onc.2014.1

Buscarlet, M., Hermann, R., Lo, R., Tang, Y., Joachim, K., & Stifani, S. (2009). Cofactor-Activated Phosphorylation Is Required for Inhibition of Cortical Neuron Differentiation by Groucho/TLE1. *PLoS ONE*, *4*, 14. doi:10.1371/journal.pone.0008107

Buscarlet, M., & Stifani, S. (2007). The "Marx" of Groucho on development and disease. *Trends in Cell Biology*, *17*(7), 353–361. doi:10.1016/j.tcb.2007.07.002

Calfon, M., Zeng, H., Urano, F., Till, J. H., Hubbard, S. R., Harding, H. P., Ron, D. (2002). IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature*, *415*(6867), 92–6. doi:10.1038/415092a

Capelli, C., Redhead, N., Abernethy, J. K., Gratrix, F., Wilson, J. F., Moen, T., I-, R. (2003). A Y Chromosome Census of the British Isles, *13*, 979–984. doi:10.1016/S

Carroll, J. S., Meyer, C. a, Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J., Brown, M. (2006). Genome-wide analysis of estrogen receptor binding sites. *Nature Genetics*, *38*(11), 1289–1297. doi:10.1038/ng1901

Chen, G., Fernandez, J., Mische, S., & Courey, A. J. (1999). A functional interaction between the histone deacetylase Rpd3 and the corepressor Groucho in Drosophila development, 2218–2230.

Chodaparambil, J. V, Pate, K. T., Hepler, M. R. D., Tsai, B. P., Muthurajan, U. M., Luger, K., … Weis, W. I. (2014). Molecular functions of the TLE tetramerization domain in Wnt target gene repression. *The EMBO Journal*, 1–13. doi:10.1002/embj.201387188

Choi, C. Y., Kim, Y. H., Kwon, H. J., & Kim, Y. (1999). The Homeodomain Protein NK-3 Recruits Groucho and a Histone Deacetylase Complex to Repress Transcription *.

Chung, C. C., Kanetsky, P. a, Wang, Z., Hildebrandt, M. a T., Koster, R., Skotheim, R. I., … Nathanson, K. L. (2013). Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nature Genetics*, *45*(6), 680–5. doi:10.1038/ng.2634

Ciarapica, R., Methot, L., Tang, Y., Lo, R., Dali, R., Buscarlet, M., Stifani, S. (2014). Prolyl isomerase Pin1 and protein kinase HIPK2 cooperate to promote cortical neurogenesis by suppressing Groucho/TLE:Hes1-mediated inhibition of neuronal differentiation. *Cell Death and Differentiation*, *21*(2), 321–32. doi:10.1038/cdd.2013.160

Cooney, R., Baker, J., Brain, O., Danis, B., Pichulik, T., Allan, P., Simmons, A. (2010). NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nature Medicine*, *16*(1), 90–97. doi:10.1038/nm.2069

Courey, A. J., & Jia, S. (2001). Transcriptional repression : the long and the short of it, 2786–2796. doi:10.1101/gad.939601.and

Cui, J., Stahl, E. a., Saevarsdottir, S., Miceli, C., Diogo, D., Trynka, G., Plenge, R. M. (2013). Genome-Wide Association Study and Gene Expression Analysis Identifies CD84 as a Predictor of Response to Etanercept Therapy in Rheumatoid Arthritis. *PLoS Genetics*, *9*(3). doi:10.1371/journal.pgen.1003394

Cuthbert, A. P., Fisher, S. A., Mirza, M. M., King, K., Hampe, J., Croucher, P. J. P., … Mathew, C. G. (2002). The Contribution of NOD2 Gene Mutations to the Risk and Site of Disease in Inflammatory Bowel Disease. *Gastroenterology*, 867–874.

Dahan, S., Roda, G., Pinn, D., Roth-Walter, F., Kamalu, O., Martin, A. P., & Mayer, L. (2008). Epithelial: lamina propria lymphocyte interactions promote epithelial cell differentiation. *Gastroenterology*, *134*(1), 192–203. doi:10.1053/j.gastro.2007.10.022

Daniels, D. L., & Weis, W. I. (2005). Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nature Structural & Molecular Biology*, *12*(4), 364–71. doi:10.1038/nsmb912

Dastidar, S. G., Narayanan, S., Stifani, S., & D'Mello, S. R. (2012). Transducin-like enhancer of Split-1 (TLE1) combines with Forkhead box protein G1 (FoxG1) to promote neuronal survival. *The Journal of Biological Chemistry*, *287*(18), 14749–59. doi:10.1074/jbc.M111.328336

De La Cadena, S. G., Hernández-Fonseca, K., Camacho-Arroyo, I., & Massieu, L. (2014). Glucose deprivation induces reticulum stress by the PERK pathway and caspase-7- and calpain-mediated caspase-12 activation. *Apoptosis*, *19*(3), 414–427. doi:10.1007/s10495-013-0930-7

Dekken, H. Van, Wink, J. C., Vissers, K. J., Franken, F., Schouten, W. R., Hop, W. C. J., … Woude, C. J. Van Der. (2007). Wnt pathway-related gene expression during malignant progression in ulcerative colitis, *109*, 266–272. doi:10.1016/j.acthis.2007.02.007

Deuring, J. J., de Haar, C., Koelewijn, C. L., Kuipers, E. J., Peppelenbosch, M. P., & van der Woude, C. J. (2011). Absence of ABCG2-mediated mucosal detoxification in patients with active inflammatory bowel disease is due to impeded protein folding. *Biochemical Journal*, *441*(1), 87–93. doi:10.1042/BJ20111281

Deuring, J. J., Fuhler, G. M., Konstantinov, S. R., Peppelenbosch, M. P., Kuipers, E. J., de Haar, C., & van der Woude, C. J. (2014). Genomic ATG16L1 risk allele-restricted Paneth cell ER stress in quiescent Crohn's disease. *Gut*, *63*(7), 1081–91. doi:10.1136/gutjnl-2012-303527

Dubnicoff, T., Valentine, S. a., Chen, G., Shi, T., Lengyel, J. a., Paroush, Z., & Courey, A. J. (1997). Conversion of Dorsal from an activator to a repressor by the global corepressor Groucho. *Genes and Development*, *11*(22), 2952–2957. doi:10.1101/gad.11.22.2952

Duerr, R. H., Taylor, K. D., Brant, S. R., Rioux, J. D., Silverberg, M. S., Daly, M. J., … Cho, J. H. (2006). A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. *Science*, *314*(5804), 1461–1463. doi:10.1126/science.1135245

Eberhard, D., Jime, G., Heavey, B., & Busslinger, M. (2000). Transcriptional repression by Pax5 (BSAP) through interaction with corepressors of the Groucho family, *19*(10).

Economou, M., Trikalinos, T. a, Loizou, K. T., Tsianos, E. V, & Ioannidis, J. P. a. (2004). Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *The American Journal of Gastroenterology*, *99*(12), 2393–404. doi:10.1111/j.1572-0241.2004.40304.x

Fisher, A. L., Ohsako, S., & Caudy, M. (1996). The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain . The WRPW Motif of the Hairy-Related Basic Helix-Loop-Helix Repressor Proteins Acts as a. *Microbiology*.

Foo, W. C., Cruise, M. W., Wick, M. R., & Hornick, J. L. (2011). Immunohistochemical staining for TLE1 distinguishes synovial sarcoma from histologic mimics. *American Journal of Clinical Pathology*, *135*(6), 839–44. doi:10.1309/AJCP45SSNAOPXYXU

Fraga, M. F., Berdasco, M., Ballestar, E., Ropero, S., Lopez-Nieva, P., Lopez-Serra, L., … Esteller, M. (2008). Epigenetic Inactivation of the Groucho Homologue Gene TLE1 in Hematologic Malignancies. *Cancer Research*, *68*(11), 4116–4122. doi:10.1158/0008-5472.CAN-08-0085

Fraga, M. F., Berdasco, M., Ballestar, E., Ropero, S., Lopez-Nieva, P., Lopez-Serra, L., … Esteller, M. (2008a). Epigenetic inactivation of the Groucho homologue gene TLE1 in hematologic malignancies. *Cancer Research*, *68*(11), 4116–22. doi:10.1158/0008-5472.CAN-08-0085

Fraga, M. F., Berdasco, M., Ballestar, E., Ropero, S., Lopez-Nieva, P., Lopez-Serra, L., … Esteller, M. (2008b). Epigenetic inactivation of the Groucho homologue

gene TLE1 in hematologic malignancies. *Cancer Research*, *68*(11), 4116–22. doi:10.1158/0008-5472.CAN-08-0085

Franke, A., McGovern, D. P. B., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., … Parkes, M. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature Genetics*, *42*(12), 1118–25. doi:10.1038/ng.717

Fujino, S., Andoh, a, Bamba, S., Ogawa, a, Hata, K., Araki, Y., Fujiyama, Y. (2003). Increased expression of interleukin 17 in inflammatory bowel disease. *Gut*, *52*(1), 65–70. doi:10.1136/gut.52.1.65

Gersemann, M., Becker, S., Kübler, I., Koslowski, M., Wang, G., Herrlinger, K. R., … Stange, E. F. (2009). Differences in goblet cell differentiation between Crohn's disease and ulcerative colitis. *Differentiation; Research in Biological Diversity*, *77*(1), 84–94. doi:10.1016/j.diff.2008.09.008

Ghosh, H. S., Spencer, J. V, Ng, B., McBurney, M. W., & Robbins, P. D. (2007a). Sirt1 interacts with transducin-like enhancer of split-1 to inhibit nuclear factor kappaB-mediated transcription. *The Biochemical Journal*, *408*(1), 105–11. doi:10.1042/BJ20070817

Ghosh, H. S., Spencer, J. V, Ng, B., McBurney, M. W., & Robbins, P. D. (2007b). Sirt1 interacts with transducin-like enhancer of split-1 to inhibit nuclear factor κB-mediated transcription. *The Biochemical Journal*, *408*, 105–111. doi:10.1042/BJ20070817

Girardin, S. E., Boneca, I. G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., … Sansonetti, P. J. (2003). Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *Journal of Biological Chemistry*, *278*(11), 8869–8872. doi:10.1074/jbc.C200651200

Godman, C. a, Joshi, R., Tierney, B. R., Greenspan, E., Rasmussen, T. P., Wang, H.-W., … Giardina, C. (2008). HDAC3 impacts multiple oncogenic pathways in colon cancer cells with effects on Wnt and vitamin D signaling. *Cancer Biology & Therapy*, *7*(10), 1570–80. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2614677&tool=pmcentrez&rendertype=abstract

Goodacre, S., Helgason, a, Nicholson, J., Southam, L., Ferguson, L., Hickey, E., Sykes, B. (2005). Genetic evidence for a family-based Scandinavian settlement of Shetland and Orkney during the Viking periods. *Heredity*, *95*(2), 129–35. doi:10.1038/sj.hdy.6800661

Grbavec, D., & Stifani, S. (1996). Molecular interaction between TLE1 and the carboxyl-terminal domain of HES-1 containing the WRPW motif. *Biochemical and Biophysical Research Communications*, *223*(3), 701–5. doi:10.1006/bbrc.1996.0959

Grimes, C. L., Ariyananda, L. D. Z., Melnyk, J. E., & O'Shea, E. K. (2012). The innate immune protein Nod2 binds directly to MDP, a bacterial cell wall fragment. *Journal of the American Chemical Society*, *134*(33), 13535–13537. doi:10.1021/ja303883c

Guderian, G., Peter, C., Wiesner, J., Sickmann, A., Schulze-Osthoff, K., Fischer, U., & Grimmler, M. (2011). RioK1, a new interactor of protein arginine methyltransferase 5 (PRMT5), competes with pICln for binding and modulates PRMT5 complex composition and substrate specificity. *Journal of Biological Chemistry*, *286*(3), 1976–1986. doi:10.1074/jbc.M110.148486

Guidi, L., Costanzo, M., Ciarniello, M., De Vitis, I., Pioli, C., Gatta, L., Gasbarrini, G. B. (2005). Increased levels of NF-kappaB inhibitors (IkappaBalpha and IkappaBgamma) in the intestinal mucosa of Crohn's disease patients during infliximab treatment. *International Journal of Immunopathology and Pharmacology*, *18*(1), 155–164.

Gutierrez, O., Pipaon, C., Inohara, N., Fontalba, A., Ogura, Y., Prosper, F., … Fernandez-Luna, J. L. (2002). Induction of Nod2 in myelomonocytic and intestinal epithelial cells via nuclear factor-kappa B activation. *The Journal of Biological Chemistry*, *277*(44), 41701–5. doi:10.1074/jbc.M206473200

Halfvarson, J., & et al. (2003). Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology*, *124*(7), 1767–1773. doi:10.1016/S0016-5085(03)00385-8

Hamidov, Z., Altendorf-Hofmann, A., Chen, Y., Settmacher, U., Petersen, I., & Knösel, T. (2011). Reduced expression of desmocollin 2 is an independent prognostic biomarker for shorter patients survival in pancreatic ductal adenocarcinoma. *Journal of Clinical Pathology*, *64*(11), 990–4. doi:10.1136/jclinpath-2011-200099

Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., Schreiber, S. (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genetics*, *39*(2), 207–211. doi:10.1038/ng1954

Hasson, P., Egoz, N., Winkler, C., Volohonsky, G., Jia, S., Dinur, T., Paroush, Z. (2005). EGFR signaling attenuates Groucho-dependent repression to antagonize Notch transcriptional output. *Nature Genetics*, *37*(1), 101–5. doi:10.1038/ng1486

Henderson, P., Hansen, R., Cameron, F. L., Gerasimidis, K., Rogers, P., Bisset, W. M., Wilson, D. C. (2012). Rising incidence of pediatric inflammatory bowel disease in Scotland. *Inflammatory Bowel Diseases*, *18*(6), 999–1005. doi:10.1002/ibd.21797

Hentschke, M., & Borgmeyer, U. (2003). Identification of PNRC2 and TLE1 as activation function-1 cofactors of the orphan nuclear receptor ERRγ. *Biochemical and Biophysical Research Communications*, *312*(4), 975–982. doi:http://dx.doi.org/10.1016/j.bbrc.2003.11.025

Herrinton, L. J., Liu, L., Levin, T. R., Allison, J. E., Lewis, J. D., & Velayos, F. (2012). Incidence and Mortality of Colorectal Adenocarcinoma in Persons with Inflammatory Bowel Disease from 1998 to 2010. *Gastroenterology*, 1–8. doi:10.1053/j.gastro.2012.04.054

Hinds, D. a, McMahon, G., Kiefer, A. K., Do, C. B., Eriksson, N., Evans, D. M., Tung, J. Y. (2013). A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nature Genetics*, *45*(8), 907–11. doi:10.1038/ng.2686

Homer, C. R., Richmond, A. L., Rebert, N. A., Achkar, J.-P., & McDonald, C. (2010). ATG16L1 and NOD2 interact in an autophagy-dependent, anti-bacterial pathway implicated in Crohn's disease pathogenesis. *Gastroenterology*, *139*(5), 1630–1641.e2. doi:10.1053/j.gastro.2010.07.006

Honzawa, Y., Nakase, H., Shiokawa, M., Yoshino, T., Imaeda, H., Matsuura, M., … Chiba, T. (2014). Involvement of interleukin-17A-induced expression of heat shock protein 47 in intestinal fibrosis in Crohn's disease. *Gut*, 1902–1912. doi:10.1136/gutjnl-2013-305632

Horak, P., Tomasich, E., Vaňhara, P., Kratochvílová, K., Anees, M., Marhold, M., Krainer, M. (2014). TUSC3 loss alters the ER stress response and accelerates prostate cancer growth in vivo. *Scientific Reports*, *4*, 3739. doi:10.1038/srep03739

Hosokawa, N., Hara, T., Kaizuka, T., Kishi, C., Takamura, A., Miura, Y., Mizushima, N. (2009). Nutrient-dependent mTORC1 Association with the ULK1–Atg13–FIP200 Complex Required for Autophagy. *Molecular Biology of the Cell*, *20*(7), 1981–1991. doi:10.1091/mbc.E08-12-1248

Hu, J., & Peter, I. (2013a). Evidence of expression variation and allelic imbalance in Crohn's disease susceptibility genes NOD2 and ATG16L1 in human dendritic cells. *Gene*, *527*(2), 496–502. doi:10.1016/j.gene.2013.06.066

Hu, J., & Peter, I. (2013b). Evidence of expression variation and allelic imbalance in Crohn's disease susceptibility genes NOD2 and ATG16L1 in human dendritic cells. *Gene*, *527*(2), 496–502. doi:10.1016/j.gene.2013.06.066

Hugot et al. (2007). Prevalence of CARD15/NOD2 mutations in Caucasian healthy people. *The American Journal of Gastroenterology*, *102*(6), 1259–67. doi:10.1111/j.1572-0241.2007.01149.x

Hugot, J., Chamaillard, M., Zouali, H., Lesage, S., Ce, J., Macrykk, J., & Thomas, G. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn ' s disease. *Nature*, 599–603.

Imboden, M., Bouzigon, E., Curjuric, I., Ramasamy, A., Kumar, A., Hancock, D. B., Probst-Hensch, N. M. (2012). Genome-wide association study of lung function decline in adults with and without asthma. *Journal of Allergy and Clinical Immunology*, *129*(5), 1218–1228. doi:10.1016/j.jaci.2012.01.074

Inohara, N., Ogura, Y., Fontalba, A., Gutierrez, O., Pons, F., Crespo, J., Nuñez, G. (2003). Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *The Journal of Biological Chemistry*, *278*(8), 5509–12. doi:10.1074/jbc.C200673200

Ioachim, E. E., Michael, M. C., & Agnantis, N. J. (2004). Immunohistochemical expression of Cyclin D1 , in inflammatory bowel disease : correlation with other cell-cycle-related proteins (Rb , p53 , ki-67 and PCNA) and clinicopathological features, 325–333. doi:10.1007/s00384-003-0571-3

Iqbal, J., Dai, K., Seimon, T., Jungreis, R., Oyadomari, M., Ron, D., Hussain, M. M. (2009). NIH Public Access. *Cell*, *7*(5), 445–455. doi:10.1016/j.cmet.2008.03.005.IRE1

Jagdis, A., Rubin, B. P., Tubbs, R. R., Pacheco, M., & Nielsen, T. O. (2009). Prospective evaluation of TLE1 as a diagnostic immunohistochemical marker in synovial sarcoma. *The American Journal of Surgical Pathology*, *33*(12), 1743–51. doi:10.1097/PAS.0b013e3181b7ed36

Jin, L.-H., Shao, Q.-J., Luo, W., Ye, Z.-Y., Li, Q., & Lin, S.-C. (2003). Detection of point mutations of the Axin1 gene in colorectal cancers. *International Journal of Cancer. Journal International Du Cancer*, *107*(5), 696–9. doi:10.1002/ijc.11435

John, G., Hegarty, J. P., Yu, W., Berg, A., Pastor, D. M., Kelly, A. a., Lin, Z. (2011). NKX2-3 variant rs11190140 is associated with IBD and alters binding of NFAT. *Molecular Genetics and Metabolism*, *104*(1-2), 174–179. doi:10.1016/j.ymgme.2011.06.023

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., Cho, J. H. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, *491*(7422), 119–124.

Kanaan, Z., Young, M., Colliver, D., Crawford, N., Cobbs, A., Hein, D. W., & Galandiuk, S. (2011). NIH Public Access, *25*(1), 27–31.

Kaser, A., Lee, A.-H., Franke, A., Glickman, J. N., Zeissig, S., Tilg, H., Blumberg, R. S. (2008). XBP1 links ER stress to intestinal inflammation and confers

genetic risk for human inflammatory bowel disease. *Cell*, *134*(5), 743–56. doi:10.1016/j.cell.2008.07.021

Kim, E. R., & Chang, D. K. (2014). Colorectal cancer in inflammatory bowel disease: the risk, pathogenesis, prevention and diagnosis. *World Journal of Gastroenterology : WJG*, *20*(29), 9872–81. doi:10.3748/wjg.v20.i29.9872

Kim, Y.-G., Shaw, M. H., Warner, N., Park, J.-H., Chen, F., Ogura, Y., & Núñez, G. (2011). Cutting Edge: Crohn's Disease-Associated Nod2 Mutation Limits Production of Proinflammatory Cytokines To Protect the Host from Enterococcus faecalis-Induced Lethality. *Journal of Immunology (Baltimore, Md. : 1950)*, *187*(6), 2849–2852. doi:10.4049/jimmunol.1001854

Kimchi-Sarfaty, C., Oh, J. M., Kim, I.-W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V, & Gottesman, M. M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science (New York, N.Y.)*, *315*(5811), 525–8. doi:10.1126/science.1135308

Kitajima, S., Takuma, S., & Morimoto, M. (1999). Changes in Colonic Mucosal Permeability in Mouse Colitis Induced with Dextran Sulfate Sodium. *Experimental Animals*, *48*(3), 137–143. doi:10.1538/expanim.48.137

Kleinschek, M. a, Boniface, K., Sadekova, S., Grein, J., Murphy, E. E., Turner, S. P., Kastelein, R. a. (2009). Circulating and gut-resident human Th17 cells express CD161 and promote intestinal inflammation. *The Journal of Experimental Medicine*, *206*(3), 525–534. doi:10.1084/jem.20081712

Knösel, T., Chen, Y., Hotovy, S., Settmacher, U., Altendorf-Hofmann, A., & Petersen, I. (2012). Loss of desmocollin 1-3 and homeobox genes PITX1 and CDX2 are associated with tumor progression and survival in colorectal carcinoma. *International Journal of Colorectal Disease*, *27*(11), 1391–9. doi:10.1007/s00384-012-1460-4

Knösel, T., Heretsch, S., Altendorf-Hofmann, A., Richter, P., Katenkamp, K., Katenkamp, D., Petersen, I. (2010). TLE1 is a robust diagnostic biomarker for synovial sarcomas and correlates with t(X;18): analysis of 319 cases. *European Journal of Cancer (Oxford, England : 1990)*, *46*(6), 1170–6. doi:10.1016/j.ejca.2010.01.032

Kobayashi, K., Inohara, N., Hernandez, L. D., Galán, J. E., Núñez, G., Janeway, C. a, Flavell, R. a. (2002). RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems. *Nature*, *416*(6877), 194–199. doi:10.1038/416194a

Kobayashi, K. S., Chamaillard, M., & Ogura, Y. (2005). Nod2-Dependent Regulation of Innate and Adaptive Immunity in the Intestinal Tract.

Korinek, V. (1997). Constitutive Transcriptional Activation by a beta -Catenin-Tcf Complex in APC-/- Colon Carcinoma. *Science*, *275*(5307), 1784–1787. doi:10.1126/science.275.5307.1784

Kosemehmetoglu, K., Vrana, J. a, & Folpe, A. L. (2009). TLE1 expression is not specific for synovial sarcoma: a whole section study of 163 soft tissue and bone neoplasms. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc*, *22*(7), 872–8. doi:10.1038/modpathol.2009.47

Kratochvílová, K., Horak, P., Ešner, M., Souček, K., Pils, D., Anees, M., … Vaňhara, P. (2015). Tumor suppressor candidate 3 (TUSC3) prevents the epithelial-to-mesenchymal transition and inhibits tumor growth by modulating the endoplasmic reticulum stress response in ovarian cancer cells. *International Journal of Cancer*, *3*, n/a–n/a. doi:10.1002/ijc.29502

Kriegel, C., & Amiji, M. M. (2011). Dual TNF-α/Cyclin D1 Gene Silencing With an Oral Polymeric Microparticle System as a Novel Strategy for the Treatment of Inflammatory Bowel Disease. *Clinical and Translational Gastroenterology*, *2*(3), e2. doi:10.1038/ctg.2011.1

Krings, M., Capelli, C., Tschentscher, F., Geisert, H., Meyer, S., Haeseler, A. Von, Thibodeau, S. N. (2000). Mutations in AXIN2 cause colorectal cancer with defective mismatch repair, *26*(october), 146–147.

Kufer, T. a., Kremmer, E., Banks, D. J., & Philpott, D. J. (2006). Role for erbin in bacterial activation of Nod2. *Infection and Immunity*, *74*(6), 3115–3124. doi:10.1128/IAI.00035-06

Lee, I. H., Cao, L., Mostoslavsky, R., Lombard, D. B., Liu, J., Bruns, N. E., … Finkel, T. (2008). A role for the NAD-dependent deacetylase Sirt1 in the regulation of autophagy. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(9), 3374–3379. doi:10.1073/pnas.0712145105

Lee, K.-H., Biswas, a., Liu, Y.-J., & Kobayashi, K. S. (2012). Proteasomal degradation of Nod2 mediates tolerance to bacterial cell wall components. *Journal of Biological Chemistry*, *287*(47), 39800–39811. doi:10.1074/jbc.M112.410027

Levanon, D., Goldstein, R. E., Bernstein, Y., Tang, H., Goldenberg, D., Stifani, S., … Groner, Y. (1998). Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(20), 11590–5. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=21685&tool=pmcentrez&rendertype=abstract

Li, Y., Xu, S., Giles, A., Nakamura, K., Lee, J. W., Hou, X., … Zang, M. (2011). Hepatic overexpression of SIRT1 in mice attenuates endoplasmic reticulum stress and insulin resistance in the liver. *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *25*(5), 1664–1679. doi:10.1096/fj.10-173492

Lin, S. M., Du, P., Huber, W., & Kibbe, W. a. (2008). Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Research*, *36*(2), 1–9. doi:10.1093/nar/gkm1075

Lindsley, D. L., & Grell, E. H. (1968). *Genetic variations of Drosophila melanogaster*. *Carnegie Institute of Washington Publication* (Vol. 627).

Lino-Silva, L. S., Flores-Gutiérrez, J. P., Vilches-Cisneros, N., & Domínguez-Malagón, H. R. (2011). TLE1 is expressed in the majority of primary pleuropulmonary synovial sarcomas. *Virchows Archiv : An International Journal of Pathology*, *459*(6), 615–21. doi:10.1007/s00428-011-1160-4

Liu, B., Gulati, A. S., Cantillana, V., Henry, S. C., Schmidt, E. a, Daniell, X., … Taylor, G. a. (2013). Irgm1-deficient mice exhibit Paneth cell abnormalities and increased susceptibility to acute intestinal inflammation. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, *305*(8), G573–84. doi:10.1152/ajpgi.00071.2013

Liu, J., Morgan, M., Hutchison, K., & Calhoun, V. D. (2010). A study of the influence of sex on genome wide methylation. *PLoS ONE*, *5*(4). doi:10.1371/journal.pone.0010028

Liu, J. Z., van Sommeren, S., Huang, H., Ng, S. C., Alberts, R., Takahashi, A., … Weersma, R. K. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics*, (October 2014). doi:10.1038/ng.3359

Long, M. D., Martin, C., Pipkin, C. a, Herfarth, H. H., Sandler, R. S., & Kappelman, M. D. (2012). Risk of Melanoma and Non-Melanoma Skin Cancer among Patients with Inflammatory Bowel Disease. *Gastroenterology*, 1–11. doi:10.1053/j.gastro.2012.05.004

Ma, Y., & Hendershot, L. M. (2001). The unfolding tale of the unfolded protein response. *Cell*, *107*(7), 827–830. doi:10.1016/S0092-8674(01)00623-7

Margariti, A., Li, H., Chen, T., Martin, D., Vizcay-barrena, G., Alam, S., … Zeng, L. (2013). XBP1 mRNA Splicing Triggers an Autophagic Response in Endothelial Cells through BECLIN-1 Transcriptional, *288*(2), 859–872. doi:10.1074/jbc.M112.412783

Martina, J. a., Chen, Y., Gucek, M., & Puertollano, R. (2012). MTORC1 functions as a transcriptional regulator of autophagy by preventing nuclear transport of TFEB. *Autophagy*, *8*(6), 903–914. doi:10.4161/auto.19653

Martínez, a., Cuttitta, F., & Teitelman, G. (1998). Expression pattern for adrenomedullin during pancreatic development in the rat reveals a common precursor with other endocrine cell types. *Cell and Tissue Research*, *293*(1), 95–100. doi:10.1007/s004410051101

Massey, D. C. O., Bredin, F., & Parkes, M. (2008). Use of sirolimus (rapamycin) to treat refractory Crohn's disease. *Gut*, *57*(9), 1294–1296. doi:10.1136/gut.2008.157297

McLarren, K. W., Theriault, F. M., & Stifani, S. (2001). Association with the nuclear matrix and interaction with Groucho and RUNX proteins regulate the transcription repression activity of the basic helix loop helix factor Hes1. *The Journal of Biological Chemistry*, *276*(2), 1578–84. doi:10.1074/jbc.M007629200

Melhem, H., Hansmannel, F., Bressenot, a., Battaglia-Hsu, S.-F., Billioud, V., Alberto, J. M., Peyrin-Biroulet, L. (2015). Methyl-deficient diet promotes colitis and SIRT1-mediated endoplasmic reticulum stress. *Gut*, *1*, 1–12. doi:10.1136/gutjnl-2014-307030

Miceli-Richard, C., Lesage, S., Rybojad, M., Prieur, a M., Manouvrier-Hanu, S., Häfner, R., Hugot, J. P. (2001). CARD15 mutations in Blau syndrome. *Nature Genetics*, *29*(1), 19–20. doi:10.1038/ng720

Michalak, M., Parker, J. M. R., & Opas, M. (2002). Ca2+ signaling and calcium binding chaperones of the endoplasmic reticulum. *Cell Calcium*, *32*(5-6), 269–278. doi:10.1016/S0143416002001884

Millar, J. K., Christie, S., & Porteous, D. J. (2003). Yeast two-hybrid screens implicate DISC1 in brain development and function. *Biochemical and Biophysical Research Communications*, *311*(4), 1019–1025. doi:10.1016/j.bbrc.2003.10.101

Molodecky, N. a, & Kaplan, G. G. (2010). Environmental risk factors for inflammatory bowel disease. *Gastroenterology & Hepatology*, *6*(5), 339–346. doi:10.1007/s10620-014-3350-9

Molodecky, N. a, Soon, I. S., Rabi, D. M., Ghali, W. a, Ferris, M., Chernoff, G., Kaplan, G. G. (2012). Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*, *142*(1), 46–54.e42; quiz e30. doi:10.1053/j.gastro.2011.10.001

Morin, P. J. (1997). Activation of beta -Catenin-Tcf Signaling in Colon Cancer by Mutations in beta -Catenin or APC. *Science*, *275*(5307), 1787–1790. doi:10.1126/science.275.5307.1787

Murch, S. H., Braegger, C. P., Walker-Smith, J. a, & MacDonald, T. T. (1993). Location of tumour necrosis factor alpha by immunohistochemistry in chronic inflammatory bowel disease. *Gut*, *34*(12), 1705–1709. doi:10.1136/gut.34.12.1705

Nagel, A. C., Krejci, A., Tenin, G., Bravo-Patiño, A., Bray, S., Maier, D., & Preiss, A. (2005). Hairless-mediated repression of notch target genes requires the combined activity of Groucho and CtBP corepressors. *Molecular and Cellular Biology*, *25*(23), 10433–10441. doi:10.1128/MCB.25.23.10433-10441.2005

Negoro, K., Goldthorpe, S., Foxwell, B. M. J., Mathew, C. G., Forbes, A., Jewell, D. P., & Playford, R. J. (2005). Muramyl dipeptide and toll-like receptor sensitivity in NOD2 -associated Crohn ' s disease, *365*, 1794–1796.

Nigro, G., Rossi, R., Commere, P. H., Jay, P., & Sansonetti, P. J. (2014). The cytosolic bacterial peptidoglycan sensor Nod2 affords stem cell protection and links microbes to gut epithelial regeneration. *Cell Host and Microbe*, *15*(6), 792–798. doi:10.1016/j.chom.2014.05.003

Nimmo, E. R., Stevens, C., Phillips, A. M., Smith, A., Drummond, H. E., Noble, C. L., … Satsangi, J. (2011). TLE1 modifies the effects of NOD2 in the pathogenesis of Crohn's disease. *Gastroenterology*, *141*(3), 972–981.e1–2. doi:10.1053/j.gastro.2011.05.043

Noble, C. L., Abbas, A. R., Lees, C. W., Cornelius, J., Toy, K., Modrusan, Z., … Diehl, L. (2010). Characterization of intestinal gene expression profiles in Crohn's disease by genome-wide microarray analysis. *Inflammatory Bowel Diseases*, *16*(10), 1717–1728. doi:10.1002/ibd.21263

Nuthall, H., Husain, J., & McLarren, K. (2002). Role for Hes1-induced phosphorylation in Groucho-mediated transcriptional repression. *Molecular and Cellular*, *22*(2), 389–399. doi:10.1128/MCB.22.2.389

Nuthall, H. N., Joachim, K., Palaparti, A., & Stifani, S. (2002). A role for cell cycle-regulated phosphorylation in Groucho-mediated transcriptional repression. *The Journal of Biological Chemistry*, *277*(52), 51049–57. doi:10.1074/jbc.M111660200

Nuthall, H. N., Joachim, K., & Stifani, S. (2004, October). Phosphorylation of serine 239 of Groucho/TLE1 by protein kinase CK2 is important for inhibition of neuronal differentiation. *Molecular and Cellular Biology*. doi:10.1128/MCB.24.19.8395-8407.2004

Ogura et al. (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn ' s disease. *Nature*, *411*(May), 603–606.

Ogura, Y., Inohara, N., Benito, A., Chen, F. F., Yamaoka, S., & Núñez, G. (2001). Nod2, a Nod1/Apaf-1 Family Member That Is Restricted to Monocytes and Activates NF-κB. *Journal of Biological Chemistry*, *276*(7), 4812–4818. doi:10.1074/jbc.M008072200

Ogura, Y., Lala, S., Xin, W., Smith, E., Dowds, T. a, Chen, F. F., … Nuñez, G. (2003). Expression of NOD2 in Paneth cells: a possible link to Crohn's ileitis. *Gut*, *52*(11), 1591–1597. doi:10.1136/gut.52.11.1591

Okamoto, R., Tsuchiya, K., Nemoto, Y., Akiyama, J., Nakamura, T., Kanai, T., & Watanabe, M. (2009). Requirement of Notch activation during regeneration of the intestinal epithelia. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, *296*(1), G23–G35. doi:10.1152/ajpgi.90225.2008

Orholm, M., Binder, V., Rasmussen, L. P., Kyvik, K. O., & Hospital, E. (2000). Concordance of Inflammatory Bowel Disease among Danish Twins, (12).

Orholm, M., & et al. (1991). Familial occurrence of inflammatory bowel disease. *New England Journal*. Retrieved from http://www.nejm.org/doi/full/10.1056/NEJM199101103240203

Owczarek, D., Cibor, D., & Mach, T. (2010). Asymmetric dimethylarginine (ADMA), symmetric dimethylarginine (SDMA), arginine, and 8-Iso-prostaglandin F2?? (8-iso-PGF2??) level in patients with inflammatory bowel diseases. *Inflammatory Bowel Diseases*, *16*(1), 52–57. doi:10.1002/ibd.20994

Palaparti, A., Baratz, A., & Stifani, S. (1997). The Groucho / Transducin-like Enhancer of split Transcriptional Repressors Interact with the Genetically Defined Amino-terminal Silencing Domain of Histone H3 *, *272*(42), 26604–26610.

Park, B. S., Song, D. H., Kim, H. M., Choi, B.-S., Lee, H., & Lee, J.-O. (2009). The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature*, *458*(7242), 1191–1195. doi:10.1038/nature07830

Paroush, Z., Finley Jr., R. L., Kidd, T., Wainwright, S. M., Ingham, P. W., Brent, R., & Ish-Horowicz, D. (1994). Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell*, *79*(5), 805–815. doi:http://dx.doi.org/10.1016/0092-8674(94)90070-1

Peer, D., Park, E. J., Morishita, Y., Carman, C. V, & Shimaoka, M. (2008). Systemic Leukocyte-Directed siRNA Delivery Revealing Cyclin D1 as an Anti-Inflammatory Target. *Science* , *319* (5863), 627–630. doi:10.1126/science.1149859

Peter, J. P., & Muddassar, M. (2001). Association between insertion mutation in NOD2 gene and Crohn ' s disease in German and British pop ..., *357*, 1925–1928.

Pickles, L. M., Roe, S. M., Hemingway, E. J., Stifani, S., & Pearl, L. H. (2002). Crystal Structure of the C-Terminal WD40 Repeat Domain of the Human Groucho / TLE1 Transcriptional Corepressor, *10*(02), 751–761.

Polakis, P. (2012). Wnt signaling in cancer. *Cold Spring Harbor Perspectives in Biology*, *4*(5), 1837–1851. doi:10.1101/cshperspect.a008052

Prescott, N. J., Dominy, K. M., Kubo, M., Lewis, C. M., Fisher, S. a., Redon, R., … Mathew, C. G. (2010). Independent and population-specific association of risk variants at the IRGM locus with Crohn's disease. *Human Molecular Genetics*, *19*(9), 1828–1839. doi:10.1093/hmg/ddq041

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. a R., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–75. doi:10.1086/519795

Qiao, L., & Wong, B. C. Y. (2009). Role of Notch signaling in colorectal cancer. *Carcinogenesis*, *30*(12), 1979–86. doi:10.1093/carcin/bgp236

Ramasamy, S., Chen, X., Wang, J., Ding, D., & Sweetser, D. A. (2010). Abstract 3113: Loss of TLE1 tumor suppressor accelerates myeloid leukemia development in cooperation with N-Myc. *Cancer Research* , *70* (8 Supplement), 3113. doi:10.1158/1538-7445.AM10-3113

Ravikumar, B., Vacher, C., Berger, Z., Davies, J. E., Luo, S., Oroz, L. G., … Rubinsztein, D. C. (2004). Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nature Genetics*, *36*(6), 585–595. doi:10.1038/ng1362

Reimund, J., Wittersheim, C., Dumont, S., & Muller, C. (1996). Increased production of tumour necrosis factor-alpha interleukin-1 beta, and interleukin-6 by morphologically normal intestinal biopsies from patients with Crohn's. *Gut*, 684–689. Retrieved from http://gut.bmj.com/content/39/5/684.abstract

Rekhi, B., Basak, R., Desai, S. B., & Jambhekar, N. a. (2012). Immunohistochemical validation of TLE1, a novel marker, for synovial sarcomas. *The Indian Journal of Medical Research*, *136*(5), 766–75. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3573597&tool=pmcentrez&rendertype=abstract

Rhie, S. K., Coetzee, S. G., Noushmehr, H., Yan, C., Kim, J. M., Haiman, C. a., & Coetzee, G. a. (2013). Comprehensive Functional Annotation of Seventy-One Breast Cancer Risk Loci. *PLoS ONE*, *8*(5). doi:10.1371/journal.pone.0063925

Riedl, S. J., Li, W., Chao, Y., Schwarzenbacher, R., & Shi, Y. (2005). Structure of the apoptotic protease-activating factor 1 bound to ADP. *Nature*, *434*(7035), 926–933. Retrieved from http://dx.doi.org/10.1038/nature03465

Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Zhang, C. K., Boucher, G., … Mark, J. (2012). variants associated with inflammatory bowel disease, *43*(11), 1066–1073. doi:10.1038/ng.952.Deep

Rogler, G., Brand, K., Vogl, D., Page, S., Hofmeister, R., Andus, T., … Gross, V. (1998). Nuclear factor κB is activated in macrophages and epithelial cells of inflamed intestinal mucosa. *Gastroenterology*, *115*(2), 357–369. doi:10.1016/S0016-5085(98)70202-1

Rosenstiel, P., Fantini, M., Bräutigam, K., Kühbacher, T., Waetzig, G. H., Seegert, D., & Schreiber, S. (2003). TNF-α and IFN-γ regulate the expression of the NOD2 (CARD15) gene in human intestinal epithelial cells. *Gastroenterology*, *124*(4), 1001–1009. doi:10.1053/gast.2003.50157

Roth, M. P., Petersen, G. M., McElree, C., Feldman, E., & Rotter, J. I. (1989). Geographic origins of Jewish patients with inflammatory bowel disease. *Gastroenterology*, *97*(4), 900–904. Retrieved from http://ukpmc.ac.uk/abstract/MED/2777043

Rubinfeld, B. (1997). Stabilization of beta -Catenin by Genetic Defects in Melanoma Cell Lines. *Science*, *275*(5307), 1790–1792. doi:10.1126/science.275.5307.1790

Ryan, N. M., Morris, S. W., Porteous, D. J., Taylor, M. S., & Evans, K. L. (2014). SuRFing the genomics wave: an R package for prioritising SNPs by functionality. *Genome Medicine*, *6*(10), 1–13. doi:10.1186/s13073-014-0079-1

Settembre, C., Di Malta, C., Polito, V. A., Arencibia, M. G., Vetrini, F., Erdin, S., … Ballabio, A. (2011). TFEB Links Autophagy to Lysosomal Biogenesis. *Science* , *332* (6036), 1429–1433. doi:10.1126/science.1204592

Shao, D., Ni, J., Shen, Y., Liu, J., Zhou, L., Xue, H., & Huang, Y. (2015). CHOP mediates XBP1S-induced renal mesangial cell necrosis following high glucose treatment. *European Journal of Pharmacology*, *758*, 89–96. doi:10.1016/j.ejphar.2015.03.069

Simms, L. a, Doecke, J. D., Walsh, M. D., Huang, N., Fowler, E. V, & Radford-Smith, G. L. (2008). Reduced alpha-defensin expression is associated with inflammation and not NOD2 mutation status in ileal Crohn's disease. *Gut*, *57*(7), 903–10. doi:10.1136/gut.2007.142588

Singh, S. B., Davis, A. S., Taylor, G. a, & Deretic, V. (2006). Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science (New York, N.Y.)*, *313*(5792), 1438–1441. doi:10.1126/science.1129577

Song, H., Hasson, P., Paroush, Z., & Courey, A. J. (2004). Groucho Oligomerization Is Required for Repression In Vivo, *24*(10), 4341–4350. doi:10.1128/MCB.24.10.4341

Stevens, C., Henderson, P., Nimmo, E. R., Soares, D. C., Dogan, B., Simpson, K. W., … Satsangi, J. (2013). The intermediate filament protein, vimentin, is a regulator of NOD2 activity. *Gut*, *62*(5), 695–707. doi:10.1136/gutjnl-2011-301775

Stifani, S., Blaumueller, C. M., Redhead, N. J., Hill, R. E., & Artavanis-Tsakonas, S. (1992). Human homologs of a Drosophila Enhancer of Split gene product define a novel family of nuclear proteins. *Nat Genet*, *2*(2), 119–127.

Stifani, Blaumueller, Redhead, H. & A.-T. (1992). Human homologs of Drosophila Enhancer of Split gene product define a novel family of nuclear proteins. *Nature Genetics*.

Strober, W., Murray, P. J., Kitani, A., & Watanabe, T. (2006). Signalling pathways and molecular interactions of NOD1 and NOD2, *6*(January), 9–20. doi:10.1038/nri1747

Tanabe, T., Chamaillard, M., Ogura, Y., Zhu, L., Qiu, S., Masumoto, J., … Núñez, G. (2004). Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition. *The EMBO Journal*, *23*(7), 1587–1597. doi:10.1038/sj.emboj.7600175

Tardif, K. D., Mori, K., & Siddiqui, A. (2002). Hepatitis C Virus Subgenomic Replicons Induce Endoplasmic Reticulum Stress Activating an Intracellular Signaling Pathway Hepatitis C Virus Subgenomic Replicons Induce Endoplasmic Reticulum Stress Activating an Intracellular Signaling Pathway, *76*(15), 7453–7459. doi:10.1128/JVI.76.15.7453

Taylor, P. (2006). nd es io s en ce no t u, (April 2015), 37–41. doi:10.4161/cc.5.2.2306

Tetsuka, T., Uranishi, H., Imai, H., Ono, T., Sonta, S. I., Takahashi, N., … Okamoto, T. (2000). Inhibition of nuclear factor-??B-mediated transcription by association with the amino-terminal enhancer of split, a Groucho-related protein lacking WD40 repeats. *Journal of Biological Chemistry*, *275*(6), 4383–4390. doi:10.1074/jbc.275.6.4383

Thompson, N. P., Driscoll, R., Pounder, R. E., & Wakefield, a J. (1996). Genetics versus environment in inflammatory bowel disease: results of a British twin study. *BMJ (Clinical Research Ed.)*, *312*(7023), 95–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2349773&tool=pmcentrez&rendertype=abstract

Travassos, L. H., Carneiro, L. A. M., Ramjeet, M., Hussey, S., Kim, Y.-G., Magalhaes, J. G., … Philpott, D. J. (2010). Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nat Immunol*, *11*(1), 55–62. Retrieved from http://dx.doi.org/10.1038/ni.1823

Tysk, C., & Lindberg, E. (1988). Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut*, 990–996. Retrieved from http://gut.bmj.com/content/29/7/990.abstract

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, a., … Ponten, F. (2015). Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419–1260419. doi:10.1126/science.1260419

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115. doi:10.1093/nar/gks596

Uthoff, S. M. S., EICHENBERGER, M. R., LEWIS, R. K., FOX, M. P., HAMILTON, C. J., MCAULIFFE, T. L., … GALANDIUK, S. (2001). Identification of candidate genes in ulcerative colitis and Crohn's disease using cDNA array technology. *International Journal of Oncology*, *19*(4), 803–810. Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=14112422

Van Limbergen, J., Radford-Smith, G., & Satsangi, J. (2014). Advances in IBD genetics. *Nature Reviews. Gastroenterology & Hepatology*, *11*(6), 372–85. doi:10.1038/nrgastro.2014.27

Van Limbergen, J., Wilson, D. C., & Satsangi, J. (2009). The genetics of Crohn's disease. *Annual Review of Genomics and Human Genetics*, *10*, 89–116. doi:10.1146/annurev-genom-082908-150013

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, *3*(7), RESEARCH0034. doi:10.1186/gb-2002-3-7-research0034

Vandussen, K. L., Liu, T. C., Li, D., Towfic, F., Modiano, N., Winter, R., … Stappenbeck, T. S. (2014). Genetic variants synthesize to produce paneth cell phenotypes that define subtypes of crohn's disease. *Gastroenterology*, *146*(1), 200–209. doi:10.1053/j.gastro.2013.09.048

Vaňhara, P., Horak, P., Pils, D., Anees, M., Petz, M., Gregor, W., … Krainer, M. (2013). Loss of the oligosaccharyl transferase subunit TUSC3 promotes proliferation and migration of ovarian cancer cells. *International Journal of Oncology*, *42*(4), 1383–1389. doi:10.3892/ijo.2013.1824

Vanrobays, E., Gelugne, J., Gleizes, P., & Caizergues-ferrer, M. (2003). Late Cytoplasmic Maturation of the Small Ribosomal Subunit Requires RIO Proteins in Saccharomyces cerevisiae Late Cytoplasmic Maturation of the Small Ribosomal Subunit Requires RIO Proteins in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *23*(6), 2083–2095. doi:10.1128/MCB.23.6.2083

Wang, F.-M., Chen, Y.-J., & Ouyang, H.-J. (2011). Regulation of unfolded protein response modulator XBP1s by acetylation and deacetylation. *The Biochemical Journal*, *433*(1), 245–52. doi:10.1042/BJ20101293

Ward, L. D., & Kellis, M. (2012a). HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Research*, *40*(D1), 1–5. doi:10.1093/nar/gkr917

Ward, L. D., & Kellis, M. (2012b). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Research*, *40*(Database issue), D930–4. doi:10.1093/nar/gkr917

Wehkamp, J., Salzman, N. H., Porter, E., Nuding, S., Weichenthal, M., Petras, R. E., … Bevins, C. L. (2005). Reduced Paneth cell ␣ -defensins in ileal Crohn ' s disease. *Analysis*, *102*(50), 18129–18134.

Werner, L., Berndt, U., Paclik, D., Danese, S., Schirbel, a., & Sturm, a. (2012). TNF inhibitors restrict T cell activation and cycling via Notch-1 signalling in inflammatory bowel disease. *Gut*, *61*(7), 1016–1027. doi:10.1136/gutjnl-2011-301267

Widmann, B., Wandrey, F., Badertscher, L., Wyler, E., Pfannstiel, J., Zemp, I., & Kutay, U. (2012). The kinase activity of human Rio1 is required for final steps of cytoplasmic maturation of 40S subunits. *Molecular Biology of the Cell*, *23*(1), 22–35. doi:10.1091/mbc.E11-07-0639

Wilson, N. J., Boniface, K., Chan, J. R., McKenzie, B. S., Blumenschein, W. M., Mattson, J. D., … de Waal Malefyt, R. (2007). Development, cytokine profile and function of human interleukin 17-producing helper T cells. *Nature Immunology*, *8*(9), 950–957. doi:10.1038/ni1497

Wu, J., & Kaufman, R. J. (2006). From acute ER stress to physiological roles of the Unfolded Protein Response. *Cell Death and Differentiation*, *13*(3), 374–84. doi:10.1038/sj.cdd.4401840

Yang, H., McElree, C., Roth, M. P., Shanahan, F., Targan, S. R., & Rotter, J. I. (1993). Familial empirical risks for inflammatory bowel disease: differences between Jews and non-Jews. *Gut*, *34*(4), 517–24. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1374314&tool=pmcentrez&rendertype=abstract

Yang, Y., & Bedford, M. T. (2013). Protein arginine methyltransferases and cancer. *Nature Reviews. Cancer*, *13*(1), 37–50. doi:10.1038/nrc3409

Zakrzewicz, D., Zakrzewicz, A., Preissner, K. T., Markart, P., & Wygrecka, M. (2012). Protein arginine methyltransferases (PRMTs): Promising targets for the treatment of pulmonary disorders. *International Journal of Molecular Sciences*, *13*(10), 12383–12400. doi:10.3390/ijms131012383

Zhao, Y., Li, X., Cai, M., Ma, K., Yang, J., Zhou, J., … Zhu, W. (2013). XBP-1u suppresses autophagy by promoting the degradation of FoxO1 in cancer cells, 491–507. doi:10.1038/cr.2013.2

Zheng, W., Rosenstiel, P., Huse, K., Sina, C., Valentonyte, R., Mah, N., … Hampe, J. (2006). Evaluation of AGR2 and AGR3 as candidate genes for inflammatory bowel disease. *Genes and Immunity*, *7*(1), 11–18. doi:10.1038/sj.gene.6364263

Zheng, X., Tsuchiya, K., Okamoto, R., Iwasaki, M., Kano, Y., Sakamoto, N., … Watanabe, M. (2011). Suppression of hath1 gene expression directly regulated by hes1 via notch signaling is associated with goblet cell depletion in ulcerative colitis. *Inflammatory Bowel Diseases*, *17*(11), 2251–2260. doi:10.1002/ibd.21611

Zurek, B., Schoultz, I., Neerincx, A., Napolitano, L. M., Birkner, K., Bennek, E., … Kufer, T. a. (2012). TRIM27 negatively regulates NOD2 by ubiquitination and proteasomal degradation. *PLoS ONE*, *7*(7). doi:10.1371/journal.pone.0041255